A TEXT-BOOK OF

MATHEMATICS

VOL. II

M. K. SINGAL & A. R. SINGAL

PITAMBAR PUBLISHING COMPANY

A TEXT-BOOK OF MATHEMATICS VOL. II

SRINIVASA RAMANUJAN (1887-1920)

Srinivasa Ramanujan, the greatest mathematical genius produced in India, was born on the 22nd December, 1887 in Tamil Nadu. He belonged to a poor Brahmin family. He earned a name for extra-ordinary mathematical ability even as a child. In 1913, Ramanujan joined the University of Madras as the first research scholar of the University. In 1914, he went to England where he collaborated with Hardy and Littlewood to produce some of the most outstanding work. In 1918, he was elected a Fellow of the Royal Society. In 1917, Ramanujan fell ill in England and returned back to Madras in 1919. He passed away on the 26th April, 1920. Even on his deathbed, he produced research work of the highest order. Ramanujan used to write on notebooks. His notebooks contain more than three thousand important theorems.

Ramanujan will be remembered not only because his work has kept first rate mathematicians busy for nearly seventy years even after his death, but also because he was able to do so without any formal training and without any means of support.

Based on the latest syllabus in Mathematics for Class XII of Senior Secondary Examination conducted by the Central Board of Secondary Education, New Delhi under the All India and Delhi Schemes.

A TEXT-BOOK OF MATHEMATICS

VOL. II

(FOR CLASS XII)

By

M.K. SINGAL
M.A., Ph.D., F.N.A.Sc., F.I.M.A. (U.K.)
Professor and Head, Department of Mathematics
Meerut University, Meerut

and

MRS. ASHA RANI SINGAL
M.A., Ph.D., F.N.A.Sc., F.I.M.A. (U.K.)
Professor of Mathematics
Meerut University, Meerut

PITAMBAR PUBLISHING COMPANY
Educational Publishers
888, East Park Road, Karol Bagh,
New Delhi-110 005

Published by
PITAMBAR PUBLISHING COMPANY
Educational Publishers
888, East Park Road
Karol Bagh,
New Delhi-110005 (India)
Telephones Off. : 770067, 776058, 526933
Res. : 5715182, 5721321, 586788

First Edition : 1990

Price : Rs. 85/-

Code No. : 27510

Printed at :
SAINI PRINTERS
9/18, Bazar Marg
Vishwas Nagar, Shahdara
Delhi-110032
Phone : 2201564

Dedicated
To
The Future Mathematics
And
Users of Mathematics

गणितस्तुतिः

यथा शिखा मयूराणां नागानां मणयो यथा ।
तद्वद्वेदाङ्गशास्त्राणां गणितं मूर्ध्नि संस्थितम् ॥ १ ॥

॥ वेदाङ्गज्यतिषात् ॥

बहुभिर्विप्रलापैः किं त्रैलोक्ये सचराचरे ।
यत्किञ्चिद्वस्तु तत्सर्वं गणितेन विना न हि ॥ २ ॥

॥ गणितसारसङ्ग्रहात् ॥

उत्पादकं यत्प्रवदन्ति बुद्धे—
रधिष्ठितं सत्पुरुषेण साङ्ख्या ।
कृत्स्नस्य लोकस्य तदेकबीज—
मव्यक्तमीशं गणितञ्च वन्दे ॥ ३ ।

॥ भास्कराचार्यः ॥

गणितसूक्ष्मं गणितं—
सङ्कलनादि तदेव सूक्ष्मम् ।
सूक्ष्मबुद्धिगम्यत्वाच्—
छ्युते च वज्रान्तं गणितम् ॥ ४ ॥

॥ स्थानाङ्गसूत्रटीकात् ॥

GREEK ALPHABET

Alpha	Α	α
Beta	Β	β
Gamma	Γ	γ
Delta	Δ	δ
Epsilon	Ε	ε
Zeta	Ζ	ζ
Eta	Η	η
Theta	Θ	θ
Iota	Ι	ι
Kappa	Κ	κ
Lamda	Λ	λ
Mu	Μ	μ
Nu	Ν	ν
Xi	Ξ	ξ
Omicron	Ο	ο
Pi	Π	π
Rho	Ρ	ρ
Sigma	Σ	σ
Tau	Τ	τ
Upsilon	Υ	υ
Phi	Φ	φ
Chi	Χ	χ
Psi	Ψ	ψ
Omega	Ω	ω

SYMBOLS

$\sim$	negation
$\wedge$	conjunction (and)
$\vee$	disjunction (or)
$\Rightarrow$	implies
$\Leftrightarrow$	is equivalent to
$\{\ \}$	set
$\in$	is an element of
$\notin$	is not an element of
$:$	such that
$\subset$	is contained in (is a subset of)
$\supset$	contains (is a superset of)
$X \sim A$	complement of A with respect to X
$\cup$	union
$\cap$	intersection
ϕ	the empty set
$\exists$	there exists
$\forall$	for all
$\mathbf{N}$	the set of natural numbers
$\mathbf{Z}$	the set of integers
$\mathbf{Q}$	the set of rational numbers
$\mathbf{Q}^+$	the set of positive rational numbers
$\mathbf{R}$	the set of real munbers
$\mathbf{R}^+$	the set of positive real numbers
$\mathbf{C}$	the set of complex numbers
$\therefore$	therefore
$\because$	because

Hindus were the first to use symbols for the various operations and the unknowns in algebra. For example, ka (क) was used for square-root and the first letters of words for various coiours were used for the unknowns. For example, ह (for हरित), न (for नील) and so on. The method of solving a quadratic by completing the square was also given to the world by Hindus. Today the solution of polynominal equations like $ax+b=0$, $ax^2+bx+c=0$ is regarded trivial. But once upon a time when people had no symbols to write an equation, the solution of even particular linear and quadratic equations was considered a great achievment. People usually guarded the solutions and posed these as challenging problems. Hence the importance of symbolism.

★ ★ ★

As the sun eclipses the stars by its brilliance, so the man of knowledge will eclipse the fame of others in assemblies of people if he proposes algebraic problems, and still more if he solves them.

—Brahmagupta

PREFACE

The book has been specially designed as a text for use in class XII of Senior Secondary Schools (under the 10+2 pattern of education). In respect of subject matter content, it strictly covers the syllabus prescribed by the Central Board of Secondary Education, New Delhi.

In the preparation of the book, the authors have kept in view the idea of an Integrated Approach to Mathematics which has now been universally accepted as a sound pedagogical principle in Mathematics Education. Wherever possible, a new mathematical concept has been introduced in the setting of real life situations, as an abstracting model, rather than an abstraction in itself. The concepts and techniques learnt have been sought to be applied to practical problems from various co-curricular subjects like Physics, Chemistry, Biology, Economics etc. An attempt has been made to present mathematics as a single entity.

The exposition is simple, yet rigorous. The language is such as a student at this level can easily follow. Since sets provide the most convenient medium in which mathematical ideas find their simplest expression, therefore, the language of sets has been used throughout the book. A proper balance between the learning of concepts and proofs, and the mastery of skills has been sought to be achieved throughout the book.

Short biographical notes have been added at appropriate places to give the student some idea about the Makers of Mathematics. Full page photographs of such mathematical giants as Jakob Bernoulli, George Boole, Arthur Cayley, Augustin Louis Cauchy, Lenohard Euler, Ronald A. Fisher, Carl Friedrich Gauss, Josiah Willard Gibbs, L'Hopital, Joseph Louis Lagrange, G.W. Leibniz, Blaise Pascal, John von Neumann, Issac Newton, Srinivasa Ramanujan, and George Bernhard Riemann have been included in the book to add to the historical perspective and to enhance the aesthetic appeal. Historical notes have been given wherever necessary.

Throughout the book a large number of examples have been solved to illustrate the various concepts and techniques. The problems have been carefully selected and properly graded and the answers have been thoroughly checked. They have been given in the form of problem-sets at the end of each section, and their number is just the right one for having a proper understanding of

the subject as well as for acquiring the necessary computational skills. A serious effort has been made to keep the book free from mistakes.

At the end of each chapter a brief summary of the chapter, a set of objective type questions, and a set of review exercises has been given. Trigonometric and logarithmic tables have been given at the end of the book.

It is hoped that the book will be found useful by all those for whom it is meant. Suggestions for the improvement of the book will be gratefully received and acknowledged.

Meerut
April 13, 1990

M.K. SINGHAL
ASHA RANI SINGHAL

CONTENTS

ARTHUR CAYLEY (1821-1895)

Cayley was born on August 16, 1821 in Richmond, England. Early in life he developed an amazing proficiency in numerical calulations. Encouraged by his teachers, he took to mathematics, winning over the resistance put up by his father. At the early age of twenty-one, Cayley became a Senior Wrangler in mathematical tripos. For fourteen years Cayley practised law but did not divorce mathematics. After that the became a professional mathematician. Shy and reserved, the frail-looking Cayley was nevertheless a great pillar of physical as well as mental strength. Serene and enduring, his hobbies included tramping, mountaineering, water colour sketching and novel-reading. He was blessed with a unique memory. He never ever forgot anything he had seen or heard. Through severely ill, Cayley continued to create mathematics till his last day which was January 26, 1895.

From the point of view of prolific inventiveness, Cayley must be put in a class, the only two other members of which happen to be Euler and Cauchy. Cayley is best remembered for three things. The first of these is his theory of algebraic invariants which is of great importance to the physicists in the theory of relativity, and is a remarkably beautiful piece of pure mathematics. The second is his invention of the geometry of higher space (space of n dimensions). But for this invention of Cayley, Klein should not have been known for his Erlanger Programme. The third was his invention of matrices Sixty-seven years after the matrices were invented by Cayley, Heisenberg realised that they provided the algebraic tool needed by him in his revolutionary work in quantum mechanics.

CHAPTER 1

Matrices and Determinants

1·1. INTRODUCTION

Matrices form one of the most important concepts of linear algebra. They have a wide variety of applications. There are many situations when a large amount of data has to be stored for future use. This data is conveniently arranged in the form of a table with rows and columns. Whenever a problem reduces to a set of linear equations, then matrix methods are applied to solve them. Matrices have important applications to geometry, electrical network theory, probability, graph theory etc.

In the present chapter we shall study some basic concepts of matrices and apply them to solutions of systems of linear equations.

1·2. DEFINITION OF A MATRIX

Let S be any set. A set of *mn* elements of S arranged in a rectangular array of *m* rows and *n* columns as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

is called an $m \times n$ ("*m* by *n*") *matrix* over S.

Illustrations

1. $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ is a 2×3 matrix over the set **N** of natural numbers.

2. $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ is a 2×2 matrix over the set **Z** of integers.

3. $\begin{pmatrix} 1 & \sqrt{2} & \sqrt{3} \\ 0 & 1 & -1 \end{pmatrix}$ is a 2×3 matrix over the set **R** of real numbers.

4. $\begin{pmatrix} 1 & i & -1 \\ 2 & -1 & i \\ 4 & 5 & 3+i \end{pmatrix}$ is a 3×3 matrix over the set **C** of complex numbers.

A matrix may be represented by the symbols $\| a_{ij} \|$, (a_{ij}), $[a_{ij}]$ or by a single letter such as A. The a_{ij}'s in a matrix are called the elements of the matrix. The indices i and j of an element indicate respectively the row and the column in which the element a_{ij} is located. Thus in the illustration 3 above, $\sqrt{3}$ is the (1, 3)th element a_{13} situated in the first row and the third column.

Since we shall be dealing only with matrices over the set of real numbers, therefore, henceforth it will be understood that the word 'matrix' stands for 'matrix over **R**' unless stated otherwise.

The $1 \times n$ matrices are called row vectors and the $m \times 1$ matrices are called column vectors. The $m \times n$ matrix whose elements are all 0 is called the null matrix (or zero matrix) of the type $m \times n$. It is usually denoted by $O_{m \times n}$, or simply by O if there is no chance of confusion.

If $m=n$, the matrix is called a *square matrix* of order n or an n-rowed square matrix. The elements $a_{11}, a_{22}, \cdots\cdots, a_{mn}$ of a square matrix A are said to constitute the *main diagonal* of A. A square matrix in which all elements except those in the main diagonal are zero, is called a *diagonal matrix*. Thus an n-rowed square matrix $[a_{ij}]$ is a diagonal matrix iff $a_{ij}=0$ whenever $i \neq j$. An n-rowed diagonal matrix $[a_{ij}]$ is sometimes also written as

$$\text{dia. } [a_{11}, a_{22}, \cdots\cdots, a_{nn}].$$

A diagonal matrix in which all the diagonal elements are equal, is called a *scalar matrix*. Thus, an n-rowed square matrix $[a_{ij}]$ is a scalar matrix iff for some number k,

$$a_{ij} = \left[\begin{array}{l} k, \text{ when } i=j, \\ 0, \text{ when } i \neq j. \end{array} \right.$$

A scalar matrix in which each diagonal element is unity, is called a *unit matrix* or an *identity matrix*. Thus, an n-rowed square matrix $[a_{ij}]$ is a unit matrix iff

$$a_{ij} = \left[\begin{array}{l} 1, \text{ whenever, } i=j, \\ 0, \text{ whenever, } i \neq j. \end{array} \right.$$

We shall denote the n-rowed unit matrix by the symbol I_n.

The matrix of elements which remain after deleting any number of rows (of course *not all* !) and columns of a matrix A is called a *submatrix* of A. The number of rows deleted need not be the same as the number of columns deleted.

Illustrations

1. $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is the 2×3 null matrix.

2. $\begin{pmatrix} 2 & 8 & 1 \\ 4 & 3 & 9 \\ 7 & 9 & 6 \end{pmatrix}$ is a 3-rowed square matrix ; 2, 3, 6 constitute the main diagonal of this matrix.

3. $\begin{pmatrix} 3 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 6 \end{pmatrix}$ is a 3-rowed diagonal matrix.

4. $\begin{pmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{pmatrix}$ is a 3-rowed scalar matrix.

5. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is the 3-rowed unit matrix. We denote it by I_3.

6. The matrix $\begin{pmatrix} 4 \\ 2 \end{pmatrix}$ is a sub-matrix of $\begin{pmatrix} 9 & 3 & 4 \\ 7 & 4 & 3 \\ 0 & 2 & 5 \end{pmatrix}$ because it can be obtained from the latter by deleting the first row and the first and the third columns.

1·3. EQUALITY OF MATRICES

Two matrices are said to be comparable when each of them has as many rows and columns as the other. Two matrices, $A=[a_{ij}]$ and $B=[b_{ij}]$, are called equal if

(*i*) they are comparable,

(*ii*) $a_{ij}=b_{ij}$ for each pair of subscripts i and j.

Thus for example, the matrices

$$\begin{pmatrix} 1 & 2 \\ 5 & 4 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \end{pmatrix}$$

are not comparable ; the matrices

$$\begin{pmatrix} 3 & 1 & 7 \\ 8 & 9 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 3 & 1 & 7 \\ 8 & 9 & 4 \end{pmatrix}$$

are comparable but not equal ; the matrices

$$\begin{pmatrix} 3 & 1 & 7 \\ 8 & 9 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} \sqrt{9} & 1 & 7 \\ 4{\cdot}2 & 9 & 1 \end{pmatrix}$$

are equal.

From the above definition, it can be easily verified that

(*i*) If A is any matrix, then $A=A$ (*reflexivity*).

(*ii*) If $A=B$, then $B=A$ (*symmetry*).

(*iii*) If $A=B$ and $B=C$, then $A=C$ (*transitivity*).

In other words, the relation of equality in the set of all matrices is an equivalence relation.

1·4. ADDITION OF MATRICES

If A and B be two comparable matrices, their sum $A+B$ is defined to be the matrix obtained by adding the corresponding elements of A and B.

For example, if

$$A=\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},\quad B=\begin{pmatrix} 3 & 7 \\ 1 & 0 \end{pmatrix},$$

then

$$A+B=\begin{pmatrix} 1+3 & 2+7 \\ 3+1 & 4+0 \end{pmatrix}=\begin{pmatrix} 4 & 9 \\ 4 & 4 \end{pmatrix}.$$

In general, if

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},\quad B=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

then

$$A+B=\begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{pmatrix}$$

1·5. PROPERTIES OF MATRIX ADDITION

Let A, B, C be three comparable matrices, say each of type $m\times n$. Then A+(B+C) and (A+B)+C are also comparable, each being of type $m\times n$. Denoting the (i, j)th element of A, B and C by a_{ij}, b_{ij}, c_{ij} respectively and observing that

$$a_{ij}+(b_{ij}+c_{ij})=(a_{ij}+b_{ij})+c_{ij},$$

it follows that

$$A+(B+C)=(A+B)+C.$$

Hence *addition of matrices is associative.*

Similarly it can be shown that if A and B are any two comparable matrices, then

$$A+B=B+A.$$

That is, matrix addition is commutative.

If A be any $m\times n$ matrix, and O be the $m\times n$ null matrix, then

$$A+O=O+A=A.$$

Finally, if A be any $m\times n$ matrix, then we can find an $m\times n$ matrix B such that

$$A+B=B+A=O,$$

O being the $m\times n$ null matrix. In fact, if $A=[a_{ij}]$, then B is the matrix whose (i, j)th element is $-a_{ij}$. The matrix B described just now is called the additive inverse (or negative) of A and is denoted by $-A$. $A-B$ is often used to denote the sum $A+(-B)$.

1·6. MULTIPLICATION OF A MATRIX BY A SCALAR

If $A=[a_{ij}]$ be an $m\times n$ matrix and k be any real number, then kA is defined to be the $m\times n$ matrix whose (i, j)th element is ka_{ij}.

Thus for example,

if $A=\begin{pmatrix} 2 & 1 \\ 7 & 9 \end{pmatrix}$, then $3A=\begin{pmatrix} 6 & 3 \\ 21 & 27 \end{pmatrix}$.

The matrix kA is called the scalar multiple of A by k. The following properties of scalar multiplication can be easily proved.

(*i*) *If A and B are comparable matrices and k is any real number, then*

$$k(A+B)=kA+kB$$

(*ii*) *If k, l be any real numbers and A be any matrix, then*

$$(k+l)A=kA+lA.$$

(*iii*) *If k, l be any real numbers and A be any matrix then*

$$k(lA)=(kl)A.$$

(*iv*) *For each matrix* A,

$$IA=A.$$

The simple pro fs of the above properties are left as exercises for the reader.

1·7. LINEAR COMBINATIONS OF MATRICES

Consider the matrices

$$A=\begin{pmatrix} 1 & -1 & 3 \\ 2 & 4 & 0 \end{pmatrix},\ B=\begin{pmatrix} 3 & 1 & -2 \\ 7 & -2 & 3 \end{pmatrix}$$

A and B are both 2×3 matrices. Let us compute the matrix $3A-2B$.

$$3A=\begin{pmatrix} 3 & -3 & 9 \\ 6 & 12 & 0 \end{pmatrix},\ -2B=\begin{pmatrix} -6 & -2 & 4 \\ -14 & 4 & -6 \end{pmatrix}$$

$$\begin{aligned}3A-2B&=3A+(-2B)\\ &=\begin{pmatrix} 3 & -3 & 9 \\ 6 & 12 & 0 \end{pmatrix}+\begin{pmatrix} -6 & -2 & 4 \\ -14 & 4 & -6 \end{pmatrix}\\ &=\begin{pmatrix} 3-6 & -3-2 & 9+4 \\ 6-14 & 12+4 & 0-6 \end{pmatrix}\\ &=\begin{pmatrix} -3 & -5 & 13 \\ -8 & 16 & -6 \end{pmatrix}\end{aligned}$$

The martix $3A-2B$ is also a 2×3 matrix. We say that it is a linear combination of the matrices A and B. The matrices $6A+3B$, $-5A-2B$ and $A+2B$ are also linear combinations of A and B. More generally, we have the following :

Definition 1·1. *If A and B are $m\times n$ matrices, and p, q are real numbers, then $pA+qB$ is called a linear combination of the matrices A and B.*

Example 1. *Find $3A-B$ if*

$$A=\begin{pmatrix} 0 & 2 & 3 \\ 2 & 1 & 4 \end{pmatrix} \text{ and } B=\begin{pmatrix} 7 & 6 & 3 \\ 1 & 4 & 5 \end{pmatrix}$$

(*A.I.S.S.C.E.*, *1984*)

Solution.

$$3A=\begin{pmatrix} 0 & 6 & 9 \\ 6 & 3 & 12 \end{pmatrix}, \quad -B=\begin{pmatrix} -7 & -6 & -3 \\ -1 & -4 & -5 \end{pmatrix}.$$

$$3A-B=\begin{pmatrix} 0-7 & 6-6 & 9-3 \\ 6-1 & 3-4 & 12-5 \end{pmatrix},$$

$$=\begin{pmatrix} -7 & 0 & 6 \\ 5 & -1 & 7 \end{pmatrix}.$$

EXERCISE 1 (*a*)

1. For each of the following matrices A and B, find A+B.

(*a*) $A=\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix}, \quad B=\begin{pmatrix} 2 & -1 \\ 4 & 3 \end{pmatrix}.$

(*b*) $A=\begin{pmatrix} 4 & 7 & 8 \\ 3 & 1 & 9 \end{pmatrix}, \quad B=\begin{pmatrix} 6 & 2 & 1 \\ 5 & 1 & 4 \end{pmatrix}.$

2. If $A=\begin{pmatrix} 1 & 2 & 6 \\ 3 & -1 & 4 \end{pmatrix}$, write down the matrices

$3A, (-4)A, -A.$

3. If $A=\begin{pmatrix} 1 & -1 & 4 \\ 2 & 8 & 9 \end{pmatrix}$, find a matrix X such that X+A=O.

4. Find a matrix X such that

$$4X=\begin{pmatrix} 1 & 2 & 1 \\ 4 & 2 & 3 \\ -1 & 9 & 7 \end{pmatrix}.$$

5. Prove the properties of matrix addition stated in the text.

6. Evaluate

$$\cos\theta\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}+\sin\theta\begin{pmatrix} \sin\theta & -\cos\theta \\ \cos\theta & \sin\theta \end{pmatrix}.$$

7. Prove the properties of multiplication of a matrix by a scalar stated in the text.

1·8. MULTIPLICATION OF MATRICES

We shall now define the product of two matrices. Let $A=[a_{ij}]$ and $B=[b_{ij}]$ be two matrices such that B has as many rows as A has columns. For the sake of definiteness, let A be of type $m\times n$, and let B of type $n\times p$. We shall construct an $m\times p$ matrix and call it the product of A and B (denoted by AB). To determine the (i, j)th element of AB, we proceed as follows :

Multiply the $(i, 1)$th element of A with the $(1, j)$th element of B, multiply the $(i, 2)$th element of A with the $(2, j)$th element of B,

....

multiply the (i, n)th element of A with the (n, j)th element of B, and add all the products obtained above. The sum so obtained is the (i, j)th element of AB.

As an example, consider the matrices

$$A=\begin{pmatrix} 2 & -1 & 3 \\ 3 & 4 & 1 \end{pmatrix}, \quad B=\begin{pmatrix} 8 & 1 \\ -4 & 5 \\ 0 & 6 \end{pmatrix}$$

Here the number of rows in **B** (=3) is the same as the number of columns in **A**. This ensures that we can talk of AB. Let AB=$[c_{ij}]$, so that $[c_{ij}]$ is a 2×2 matrix.

Now, to write down c_{11}, we take the elements of the first row of A, namely, 2, −1, 3 (in this order) and the elements of the first column of B, namely, 8, −4, 0 (in this order). We now form the products 2. 8, (−1) (−4), 3.0 and then add them. We thus have $c_{11}=2.8+(-1)(-4)+3.0=20$.

Similarly,

$c_{12}=2.1+(-1).5+3.6=15$,

$c_{21}=3.8+4.(-4)+1.0=8$,

$c_{22}=3.1+4.5+1.6=29$.

Thus $[c_{ij}]=\begin{pmatrix} 20 & 15 \\ 8 & 29 \end{pmatrix}$ is the desired product.

We can state the above rule thus :

To obtain the (i, j)th element of the product AB, we multiply the elements of the ith row of A by the corresponding elements of the jth column of B and add the resulting products. The sum so obtained is the desired (i, j)th element of AB.

To obtain the product of two matrices A and B, it is convenient to arrange the working like this :

Write the matrices A and B as shown below (See fig. 1.1) : Now draw a pair of brackets to the right of A and below B where the product is to be written as in fig. 1.1.

Consider the point of the product brackets where the *i*th row of A and the *j*th column of B intersect. The (*i*, *j*)th element of AB has to be entered at this place.

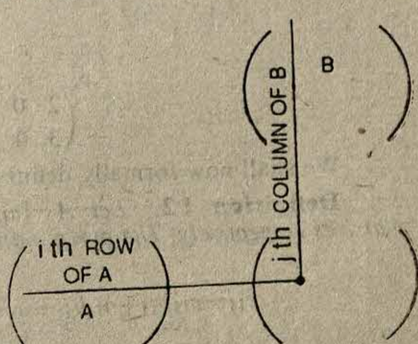


Fig. 1.1

To write down the (i, j)th element of the product, we consider the ith row of A and the jth column of B which point to it. Multiply each element of the ith row of A (starting from the left) with the corresponding element of the jth column of B (starting from the top) and add the products. For example, cnnsider the matrices

$$A=\begin{pmatrix}2 & 0 & 1\\ 3 & 4 & 1\end{pmatrix},\quad B=\begin{pmatrix}1 & 3 & 5\\ 5 & 3 & 7\\ 2 & 1 & 2\end{pmatrix}$$

To write down the product, we shall procced thus :

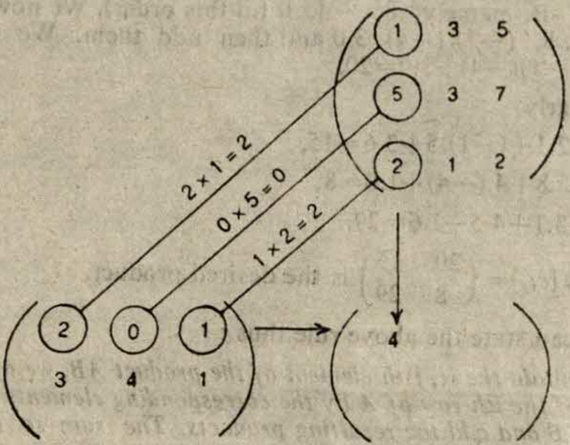


Fig. 1.2

Since $2+0+2=4$, we have put down 4 in the (1,1)th place of the product.

Proceeding in this way, it can be easily seen that the completed product is as shown below :

$$\begin{array}{cc} & \begin{pmatrix}1 & 3 & 5\\ 5 & 3 & 7\\ 2 & 1 & 2\end{pmatrix}\\ \begin{pmatrix}2 & 0 & 1\\ 3 & 4 & 1\end{pmatrix} & \begin{pmatrix}4 & 7 & 12\\ 25 & 22 & 45\end{pmatrix}.\end{array}$$

We shall now formally define the product of two matrices.

Definition 1·2. *Let $A=[a_{ij}]$, $B=[b_{ij}]$ be $m\times n$ and $n\times p$ matrices respectively. The $m\times p$ matrix $[c_{ij}]$, where*

$$c_{ij}=a_{i1}b_{1j}+a_{i2}b_{2j}+\ldots+a_{in}b_{nj}=\sum_{k=1}^{n}a_{ik}b_{kj},$$

is called the **product** *of the matrices A and B and is denoted by AB.*

The above definition says that it is possible to talk of AB only when the number of columns in A equals the numbers of rows in B. Two matrices A and B satisfying the above condition are said to be conformable to multiplication.

EXERCISE 1 (*b*)

1. Let

$$A=\begin{pmatrix}1 & 0 & 0\\ 0 & -2 & 3\\ 2 & 1 & -1\end{pmatrix},\ B=\begin{pmatrix}0 & 2 & -1\\ 3 & 1 & 0\\ 0 & -1 & 2\end{pmatrix},\ C=\begin{pmatrix}4 & 2 & 1\\ -4 & -2 & 1\\ 0 & 1 & 1\end{pmatrix}$$

Write down

(*a*) AB	(*b*) BC	(*c*) CA
(*d*) BA	(*e*) CB	(*f*) AC
(*g*) A(BC)	(*h*) B(CA)	(*i*) C(AB)
(*j*) (AB)C	(*k*) (BC)A	(*l*) (CA)B.

2. Perform the following multiplications :

(*a*) $$\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}-1 & 2 & 3\\ 2 & 0 & 1\\ 1 & 4 & 5\end{pmatrix},$$

(*b*) $$\begin{pmatrix}-3 & 5 & 7\\ 4 & 2 & 3\\ 1 & 8 & 7\end{pmatrix}\begin{pmatrix}\alpha & 0 & 0\\ 0 & \beta & 0\\ 0 & 0 & \gamma\end{pmatrix}.$$

3. If $A=\begin{pmatrix}2 & -1 & 1\\ -2 & 3 & -2\\ -4 & 4 & -3\end{pmatrix}$,

show that AA=A.

4. If $A=\begin{pmatrix}4 & -1 & -4\\ 3 & 0 & -4\\ 3 & -1 & -3\end{pmatrix}$, show that $AA=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$

5. If $A=\begin{pmatrix}-2 & 3 & -1\\ -1 & 2 & -1\\ -6 & 9 & -4\end{pmatrix}$, $B=\begin{pmatrix}1 & 3 & -1\\ 2 & 2 & -1\\ 3 & 0 & -1\end{pmatrix}$

verify that AB=BA.

6. If $U=\begin{pmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0\end{pmatrix}$, compute UU, (UU)U, ((UU)U)U.

7. If $A=\begin{pmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{pmatrix}$, $B=\begin{pmatrix}\cos\beta & -\sin\beta\\ \sin\beta & \cos\beta\end{pmatrix}$,

show that AB=BA.

8. If $A=\begin{pmatrix} -1 & 1 & -1 \\ 3 & -3 & 3 \\ 5 & -5 & 5 \end{pmatrix}$, show that $AA=A$.

9. If $\sigma_x=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_y=\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$,

$I_2=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, show that

$\sigma_x\sigma_x=\sigma_y\sigma_y=I_2$ and $\sigma_x\sigma_y=-\sigma_y\sigma_x$.

10. If $A=\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $B=\begin{pmatrix} 3 & 1 \\ 4 & 5 \end{pmatrix}$, $C=\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$, verify that

$A(BC)=(AB)C$, and $A(B+C)=AB+AC$.

11. If $A=\begin{pmatrix} -1 & 2 \\ 2 & 3 \end{pmatrix}$, $B=\begin{pmatrix} 3 & 0 \\ 1 & 1 \end{pmatrix}$, verify that

$(A+B)^2=A^2+AB+BA+B^2$.

Can this be put in the simpler form $A^2+2AB+B^2$?

12. D_1 and D_2 are two 3×3 diagonal matrices. Show that

(*i*) D_1D_2 is a diagonal matrix,

(*ii*) $D_1D_2=D_2D_1$.

1·9. PROPERTIES OF MATRIX MULTIPLICATION

We shall now consider some properties of matrix multiplication.

Consider the matrices

$$A=\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},\ B=\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

It can be easily seen that

$$AB=\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

but $$BA=\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},$$

so that $AB\neq BA$. This means that *matrix multiplication is not commutative*. In fact, for a given pair of matrices A and B, the products AB and BA may not be even comparable. For example, if A be a 2×3 matrix and B be a 3×2 matrix, then AB would be a 2×2 matrix, and BA would be a 3×3 matrix.

It can also happen that for a pair of matrices A and B, the product AB may be defined but the product BA may not be defined. For example, if A be a 2×3 matrix and B be a 3×4 matrix, then AB would be a 2×4 matrix, but it is not meaningful to talk of BA.

It may be worthwhile to note that the statement 'matrix multiplication is not commutative' does not mean that we can never have $AB=BA$. It simply means that $AB=BA$ is not always true. (That is,

there do exist some pairs of matrices A and B for which AB and BA are different.)

Consider now the matrices

$$A=\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},\ B=\begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}.$$

For these matrices, we have $AB=\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, so that AB is a zero matrix, whereas none of A and B is a zero matrix.

Thus in the context of matrices, AB=O need not always imply that either A=O or B=O. The familiar *cancellation law of multiplication for numbers fails to be true for matrix multiplication.*

Having seen that two important laws of multiplication, namely, commutativity of multiplication, and cancellation property for multiplication of numbers fail to be true for matrices, one might wonder whether any familiar property of multiplication (for numbers) will hold good for matrices. As we shall see below, some properties of multiplication for numbers do go over to matrices.

1·9·1. Associativity of Matrix Multiplication

Consider the matrices

$$A=\begin{pmatrix} 2 & 1 & 1 \\ -1 & 1 & 4 \end{pmatrix},\ B=\begin{pmatrix} 1 & 5 \\ 7 & 8 \\ -1 & 0 \end{pmatrix},\ C=\begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix}.$$

For these matrices, we have

$$A\,(BC)=\begin{pmatrix} 28 & -10 \\ 4 & -1 \end{pmatrix}=(AB)C.$$

The above equality is not a coincidence. The following theorem says that the statement A(BC)=(AB)C is always true.

Theorem 1·1. *Matrix multiplication is associative. That is, if A, B, C be of suitable sizes for the products $A(BC)$ and $(AB)C$ to exist, then $A(BC)=(AB)C$.*

Proof. Let $A=[a_{ij}]$, $B=[b_{ij}]$, $C=[c_{ij}]$ be three matrices of types $m\times n$, $n\times p$, $p\times q$ respectively. Then

$AB=[u_{ij}]$ is an $m\times p$ matrix, where $u_{ij}=\sum_{k=1}^{n} a_{ik}\, b_{kj}$.

$BC=[v_{ij}]$ is an $n\times q$ matrix, where $v_{ij}=\sum_{k=1}^{p} b_{ik}\, c_{kj}$.

(AB)C$=[w_{ij}]$ is an $m\times q$ matrix, where

$$\begin{aligned} w_{ij} &= \sum_{r=1}^{p} u_{ir}\, c_{rj}, \\ &= \sum_{r=1}^{n}\Big(\sum_{k=1}^{n} a_{ik}\, b_{kr}\Big) c_{rj}, \\ &= \sum_{k=1}^{n}\Big(a_{ik} \sum_{r=1}^{p} b_{kr}\, c_{rj}\Big), \\ &= \sum_{k=1}^{n} a_{ik}\, v_{kj}. \end{aligned}$$

Therefore, w_{ij} is also the (i, j)th element of the $m\times q$ matrix A(BC). Thus A(BC)=(AB)C.

Hence *matrix multiplication is associative*.

1·9·2. Distributive Property

Consider the matrices

$$A=\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},\ B=\begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix},\ C=\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}.$$

For these matrices

$$B+C=\begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix},\ A(B+C)=\begin{pmatrix} 1 & 0 \\ 5 & 4 \end{pmatrix},$$

$$AB=\begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix},\quad AC=\begin{pmatrix} -1 & -1 \\ 3 & 5 \end{pmatrix},$$

$$AB+AC=\begin{pmatrix} 1 & 0 \\ 5 & 4 \end{pmatrix},\ \text{so that}$$

$$A(B+C)=AB+AC.$$

This equality is only an illustration of the following theorem :

Theorem 1·2. *Multiplication of matrices is distributive with respect to addition. That is,*

$$A(B+C)=AB+AC,$$

and $$(B+C)D=BD+CD,$$

where A, B, C, D are of suitable sizes for the above equations to be meaningful.

Proof. We shall prove the first of the above statements and leave the proof of the second to the reader.

Let $A=[a_{ij}]$ be an $m\times n$ matrix and let $B=[b_{ij}]$ and $C=[c_{ij}]$ be both $n\times p$ matrices.

Since B and C are both $n \times p$ matrices, therefore, $B+C$ is an $n \times p$ matrix, and consequently $A(B+C)$ is an $m \times p$ matrix.

Again, since A is an $m \times n$ matrix and B is an $n \times p$ matrix, therefore, AB is an $m \times p$ matrix. For a similar reason, AC is also an $m \times p$ matrix. Now AB and AC being $m \times p$ matrices, $AB+AC$ is also an $m \times p$ matrix. We have thus seen that the matrices $A(B+C)$ and $AB+AC$ are both of the same size. We shall now show that their corresponding elements are equal. Let p_{kj} denote the (k, j)th element of $B+C$.

(i, j)th element of $A(B+C)$

$$=\sum_{k=1}^{n} a_{ik}\, p_{kj},$$

$$=\sum_{k=1}^{n} a_{ik}\,(b_{kj}+c_{kj}),$$

$$=\sum_{k=1}^{n} a_{ik}\, b_{kj}+\sum_{k=1}^{n} a_{ik}\, c_{kj},$$

$=(i, j)$th element of $AB+(i, j)$th element of AC,

$=(i, j)$th element of $(AB+AC)$.

Since $A(B+C)=AB+AC$ are comparable matrices with corresponding elements equal, therefore,

$$A(B+C)=AB+AC.$$

1·9·3. A Property of the Unit Matrix

Consider the matrices,

$$A=\begin{pmatrix} 1 & -3 \\ 7 & -4 \end{pmatrix},\ I_2=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

For these matrices, we have

$$AI_2=I_2A=A.$$

The matrix I_2 has thus no effect on A so far as multiplication is concerned. For this reason, I_2 is usually called the 2-rowed *unit matrix* (or *identity matrix*).

Definition 1·3. *For each positive integer n, the n-rowed square matrix*

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

where each element in the principal diagonal is 1 and every other element is 0, is called the n-rowed *unit matrix* (or *identity matrix*) and is denoted by I_n.

The unit matrices possess the following interesting and useful property :

Theorem 1·3. *If A be any* $m \times n$ *matrix, then*

$$I_m A = A = A I_n.$$

Proof. Let $A=[a_{ij}]$ be an $m \times n$ matrix. The matrix $I_m A$ is an $m \times n$ matrix and is, therefore, comparable to A.

Also, (i, j)th element of $I_m A$

$$= \sum_{k=1}^{m} (i, k)\text{th element of } I_m . a_{kj}.$$

Since the (i, k)th element of I_m is zero except when $k=i$, therefore the above sum has only one term (possibly) different from zero, namely, the ith term and this term

$$=\{(i, i)\text{th element of } I_m\}\, a_{ij}=1.a_{ij}=a_{ij}.$$

Thus (i, j)th element of $I_m A = a_{ij} = (i, j)$th element of A.

Since the matrices $I_m A$ and A are comparable and their corresponding elements are equal, therefore, we have $I_m A = A$.

Similarly it can be shown that $A = A I_n$.

Remark. Whenever there is no chance of confusion, it is usual to denote I_n simply by I.

1·10. POSITIVE INTEGRAL POWERS OF A SQUARE MATRIX

Let A be any $n \times n$ matrix. The matrix products

$$A(A(AA)),\ (A(AA))A,\ (AA)(AA),\ ((AA)A)A$$

are all meaningful. Also, as a consequence of the associative law, they are all equal. As in the case of numbers, we denote each of these products by A^4. In fact, for each positive integer k, we can define the matrix A^k. We do it as in the following inductive (or recursive) definition.

Definition 1·4. *If A be an* $n \times n$ *matrix, then*

$$A^1 = A,$$

and

$$A^{k+1} = A^k . A,$$

for each positive integer k.

Theorem 1·4. *If A be an n-rowed square matrix, then for every pair of positive integers p and q,*

$$A^p . A^q = A^{p+q}$$

and

$$(A^p)^q = A^{pq}.$$

Proof. We shall prove the first statement only and ask the reader to supply a proof for the second statement.

Since by definition, $A^p \cdot A^1 = A^{p+1}$, for each positive integer p, therefore the statement

$$A^p \cdot A^q = A^{p+q}$$

holds when $q=1$, whatever p may be.

We shall now show that if it also holds for all values of p when q has a fixed value, say k, then it also holds for all values of p, when q has the value $k+1$.

In fact, $A^p \cdot A^{k+1} = A^p \cdot (A^k \cdot A)$, by def. 1.4,
$= (A^p \cdot A^k) \cdot A$, by associativity,
$= (A^{p+k}) \cdot A$, by hypothesis,
$= A^{(p+k)+1}$, by def. 1.4,
$= A^{p+(k+1)}$, by associative law for addition of numbers.

The proof is now complete by induction.

Example 2. *If* $A = \begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix}$,

compute $A^3 + 3A^2 - 13A$. (*A.I.S.S.C.E.*, *1986*)

Solution.

$$A^2 = \begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix}$$

$$= \begin{pmatrix} 2.2+3.1 & 2.3+3(-5) \\ 1.2+(-5).1 & 1.3+(-5)(-5) \end{pmatrix}$$

$$= \begin{pmatrix} 7 & -9 \\ -3 & 28 \end{pmatrix}$$

$$A^3 = A^2A = \begin{pmatrix} 7 & -9 \\ -3 & 28 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix}$$

$$= \begin{pmatrix} 7.2+(-9).1 & 7.3+(-9)(-5) \\ -3.2+28.1 & -3.3+28(-5) \end{pmatrix}$$

$$= \begin{pmatrix} 5 & 66 \\ 22 & -149 \end{pmatrix}$$

$$A^3 + 3A^2 - 13A = \begin{pmatrix} 5 & 66 \\ 22 & -149 \end{pmatrix} + 3\begin{pmatrix} 7 & -9 \\ -3 & 28 \end{pmatrix} - 13\begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix}$$

$$= \begin{pmatrix} 5 & 66 \\ 22 & -149 \end{pmatrix} + \begin{pmatrix} 21 & -27 \\ -9 & 84 \end{pmatrix} - \begin{pmatrix} 26 & 39 \\ 13 & -65 \end{pmatrix}$$

$$= \begin{pmatrix} 5+21-26 & 66-27-39 \\ 22-9-13 & -149+84+65 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Example 3. *If B, C are n-rowed square matrices and if*

$$A = B + C,\ BC = CB,\ C^2 = 0,$$

then show that for every positive integer p,

$$A^{p+1} = B^p[B + (p+1)C].$$

Solution. We shall prove the result by induction on p. The result holds for $p=1$, for

$$\begin{aligned} A^2 &= (B+C)^2, \\ &= (B+C)(B+C), \\ &= B^2+BC+CB+C^2, \\ &= B^2+2BC, \text{ since } CB=BC,\ C^2=0, \\ &= B(B+2C). \end{aligned}$$

Let us now assume that the result holds when $p=k$. Then

$$\begin{aligned} A^{k+2} &= A^{k+1} \,.\, A, \\ &= B^k[B+(k+1)C][B+C], \\ &= B^k[B^2+(k+1)CB+BC+(k+1)C^2], \\ &= B^k[B^2+(k+2)BC], \text{ since } CB=BC,\ C^2=0. \\ &= B^{k+1}[B+(k+2)C], \end{aligned}$$

showing that the result holds when $p=k+1$.

The proof is now complete by induction.

EXERCISE 1 (*c*)

1. Let A, B, C be real 2×2 matrices, and let

$$[A, B]=AB-BA.$$

Prove that :

(*i*) $[A, A]=O$;

(*ii*) $[[A, B], C]+[[B, C], A]+[[C, A], B]=O$;

(*iii*) $[A, B]=I \Rightarrow [A, B^n]=n\, B^{n-1}$, for all positive integers n.

2. The trace of a matrix

$$A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

is defined by $T_r(A)=a_{11}+a_{22}$.

Prove that if A, B are 2×2 matrices, then

(*i*) $T_r(A+B)=T_r(A)+T_r(B)$;

(*ii*) $T_r(AB)=T_r(BA)$;

(*iii*) $T_r(I)=2$.

3. Prove, with the same notation as in problem 1, that there are no matrices satisfying $[A, B]=I$.

4. Do there exist 2×2 matrices with integer entries such that

(*a*) $AB=O,\ BA\neq O$;

(*b*) $AB=BA=O,\ A\neq O,\ B\neq O,\ A\neq B$;

(*c*) $AB=BA\neq O,\ A, B\neq I,\ A\neq B$.

In each case, if the answer is yes, then justify the answer by giving examples of suitable matrices A and B.

5. Let A denote the matrix

$$\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix},$$

and let I denote the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Prove that $A^2=3A-2I$.

Prove also, that if n is any positive integer, then

$$A^n=(2^n-1)A-2(2^{n-1}-1)I.$$

6. If $\quad A=\begin{pmatrix} p & q \\ 0 & 1 \end{pmatrix},\ B=\begin{pmatrix} 1 & q \\ 0 & 1 \end{pmatrix},$

prove by induction that

$$A^n=\begin{pmatrix} p^n & \dfrac{p^n-1}{p-1}q \\ 0 & 1 \end{pmatrix},\ \text{if } p\neq 1,$$

$$B^n=\begin{pmatrix} 1 & nq \\ 0 & 1 \end{pmatrix}.$$

7. Let A, B, C be real 2×2 matrices, and let

$$A * B=\tfrac{1}{2}(AB+BA).$$

Prove that :

(*i*) $A * B=B * A$;

(*ii*) $B * B=B^2$;

(*iii*) $A * I=A$;

(*iv*) $A * (B * C)=\tfrac{1}{4}(ABC+ACB+BCA+CBA)$;

(*v*) $A * (B+C)=(A * B)+(A * C)$:

(*vi*) $c(A * B)=(cA) * B=A * (cB)$,

where c is any real number.

8. Prove that, if

$$A=\begin{pmatrix} 0 & -\tan\alpha \\ \tan\alpha & 0 \end{pmatrix},$$

then
$$\begin{pmatrix} \cos 2\alpha & -\sin 2\alpha \\ \sin 2\alpha & \cos 2\alpha \end{pmatrix}(I-A)=I+A.$$

1·11. TRANSPOSE OF A MATRIX

Consider the matrices

$$A=\begin{pmatrix} 1 & 2 & 3 \\ -4 & 5 & 7 \end{pmatrix},\ \text{and } B=\begin{pmatrix} 1 & -4 \\ 2 & 5 \\ 3 & 7 \end{pmatrix}.$$

The matrix A is a 2×3 matrix, and the matrix B is a 3×2 matrix. Also. the first column of B is the same as the first row of A and the second column of B is the same as the second row of A. In other words, B is the matrix obtained by writing the rows of A as columns. We say that B is the transpose of A.

Definition 1·5. *If $A=[a_{ij}]$ be an $m\times n$ matrix, then the $n\times m$ matrix $B=[b_{ij}]$ such that $b_{ij}=a_{ji}$ is called the transpose of A and is denoted by A^t.*

From the above definition we find that :

(*a*) the transpose of an $m\times n$ matrix is an $n\times m$ matrix ;

(*a*) the (i, j)th element of A^t is the (j, i)th element of A.

Example 4. *Let* $A=\begin{pmatrix} 2 & -3 & 1 \\ 4 & 2 & 3 \end{pmatrix}$, $B=\begin{pmatrix} 3 & -2 & 4 \\ 1 & 3 & -5 \end{pmatrix}$.

Compute A^t, $(A^t)^t$, B^t, $(A+B)^t$, A^t+B^t, $(2A)^t$ *and* $2A^t$.

Solution.
$$A^t=\begin{pmatrix} 2 & 4 \\ -3 & 3 \\ 1 & 3 \end{pmatrix},\ (A^t)^t=\begin{pmatrix} 2 & -3 & 1 \\ 4 & 2 & 3 \end{pmatrix}.$$

$$B^t=\begin{pmatrix} 3 & 1 \\ -2 & 3 \\ 4 & -5 \end{pmatrix},$$

$$A+B=\begin{pmatrix} 5 & -5 & 5 \\ 5 & 5 & -2 \end{pmatrix},\ (A+B)^t=\begin{pmatrix} 5 & 5 \\ -5 & 5 \\ 5 & -2 \end{pmatrix},$$

$$A^t+B^t=\begin{pmatrix} 5 & 5 \\ -5 & 5 \\ 5 & -2 \end{pmatrix},\ 2A=\begin{pmatrix} 4 & -6 & 2 \\ 8 & 4 & 6 \end{pmatrix},$$

$$(2A)^t=\begin{pmatrix} 4 & 8 \\ -6 & 4 \\ 2 & 6 \end{pmatrix},\ 2A^t=2\begin{pmatrix} 2 & 4 \\ -3 & 2 \\ 1 & 3 \end{pmatrix}=\begin{pmatrix} 4 & 8 \\ -6 & 4 \\ 2 & 6 \end{pmatrix}.$$

Remark. From the above example, we find that

$$(A^t)^t=A,$$
$$(A+B)^t=A^t+B^t,$$
$$(2A)^t=2A^t.$$

The results are special cases of the following theorem :

Theorem 1·5. *If* A^t *and* B^t *be the transposes of* A *and* B *respectively, then*

(*i*) $(A^t)^t=A$.

(*ii*) $(A+B)^t=A^t+B^t$, A *and* B *being comparable.*

(*iii*) $(kA)^t=kA^t$, *k being any real number.*

Proof. (*i*) Let A be an $m\times n$ matrix. A^t, the transpose of A is an $n\times m$ matrix, and $(A^t)^t$, the transpose of A^t is an $m\times n$ matrix. The matrices $(A^t)^t$ and A are, therefore, comparable.

Also, (i, j)th element of $(A^t)^t=(j, i)$th element of A^t,
$=(i, j)$th element of A.

Thus the matrices $(A^t)^t$ and A are comparable, and their corresponding elements are equal.

Hence $(A^t)^t = A$.

(*ii*) Let A and B be $m \times n$ matrices.

Since A and B are both $m \times n$ matrices, therefore, $A+B$ exists and is an $m \times n$ matrix. Consequently $(A+B)^t$ is an $n \times m$ matrix.

Again, A^t and B^t are both $n \times m$ matrices, so that $A^t + B^t$ is also an $n \times m$ matrix. Thus the matrices $(A+B)^t$ and $A^t + B^t$ are comparable, each being of the type $n \times m$.

Also, (i, j)th element of $(A+B)^t$

$=(j, i)$th element of $A+B$,

$=(j, i)$th element of $A+(j, i)$th element of B,

$=(i, j)$th element of $A^t+(i, j)$th element of B^t,

$=(i, j)$th element of (A^t+B^t).

Thus the matrices $(A+B)^t$ and A^t+B^t are comparable, and their corresponding elements are equal.

Hence $(A+B)^t = A^t + B^t$.

(*iii*) Let A be an $m \times n$ matrix. Then kA is also an $m \times n$ matrix, so that $(kA)^t$ is an $n \times m$ matrix.

Again, A^t is an $n \times m$ matrix so that kA^t is also an $n \times m$ matrix.

Thus the matrices $(kA)^t$ and kA^t are comparable, each being of the type $n \times m$.

Also, (i, j)th element of $(kA)^t = (j, i)$th element of kA,

$=k \,.\, (j, i)$th element of A,

$=k \,.\, (i, j)$th element of A^t,

$=(i, j)$th element of (kA^t).

Thus the matrices $(kA)^t$ and kA^t are comparable and their corresponding elements are equal.

Hence $(kA)^t = kA^t$.

Remark. In view of (*i*) above, we find that if A and B be two matrices such that $B = A^t$, then $B^t = A$, *i.e.*, if B *is the transpose of* A, *then* A *is the transpose of* B.

Example 5. *If* $A=\begin{pmatrix} 2 & 4 & -1 \\ -1 & 0 & 2 \end{pmatrix}$, $B=\begin{pmatrix} 3 & 4 & 5 \\ -1 & 2 & 7 \\ 2 & 1 & 0 \end{pmatrix}$,

compute $(AB)^t$ *and* B^tA^t.

Solution. $$AB=\begin{pmatrix} 2 & 4 & -1 \\ -1 & 0 & 2 \end{pmatrix}\begin{pmatrix} 3 & 4 & 5 \\ -1 & 2 & 7 \\ 2 & 1 & 0 \end{pmatrix},$$

$$=\begin{pmatrix} 0 & 15 & 38 \\ 1 & -2 & -5 \end{pmatrix}.$$

so that $$(AB)^t=\begin{pmatrix} 0 & 1 \\ 15 & -2 \\ 38 & -5 \end{pmatrix}.$$

Also, $$B^t=\begin{pmatrix} 3 & -1 & 2 \\ 4 & 2 & 1 \\ 5 & 7 & 0 \end{pmatrix}, A^t=\begin{pmatrix} 2 & -1 \\ 4 & 0 \\ -1 & 2 \end{pmatrix},$$

so that $$B^tA^t=\begin{pmatrix} 3 & -1 & 2 \\ 4 & 2 & 1 \\ 5 & 7 & 0 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ 4 & 0 \\ -1 & 2 \end{pmatrix},$$
$$=\begin{pmatrix} 0 & 1 \\ 15 & -2 \\ 38 & -5 \end{pmatrix}.$$

Remark. From the above example we find that $(AB)^t=B^tA^t$. This result is a particular case of the following theorem, usually known as the reversal law for transposes.

Theorem 1·6. *If* A *and* B *be of suitable sizes for* AB *to exist, then* $(AB)^t=B^tA^t$.

Proof. Let $A=[a_{ij}]$, $B=[b_{ij}]$ be $m\times n$ and $n\times p$ matrices respectively.

Then $A^t=[c_{ij}]$, where $c_{ij}=a_{ji}$, is an $n\times m$ matrix,

$B^t=[d_{ij}]$, where $d_{ij}=b_{ji}$, is a $p\times n$ matrix.

The matrices $(AB)^t$ and B^tA^t are comparable, each being of the type $p\times m$.

Also, (i, j)th element of $(AB)^t=(j, i)$th element of AB,

$$=\sum_{k=1}^{n} a_{jk}b_{ki},$$

$$=\sum_{k=1}^{n} c_{kj}d_{ik}$$

$$=\sum_{k=1}^{n} d_{ik}c_{kj},$$

$=(i, j)$th element of B^tA^t.

Hence the result.

EXERCISE 1 (*d*)

1. Calculate the transpose of each of the following matrices :

$$\begin{pmatrix} 1 & -1 & 2 \\ 3 & 1 & 4 \\ -2 & 7 & 8 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 7 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 5 & 6 \\ 2 & 3 & 0 \end{pmatrix}.$$

2. For each of the following matrices A, verify that $A^t = A$:

$$\begin{pmatrix} 1 & 4 & 5 \\ 4 & 2 & 6 \\ 5 & 6 & 3 \end{pmatrix}, \begin{pmatrix} -1 & 2 & 1 \\ 2 & 0 & 7 \\ 1 & 7 & 3 \end{pmatrix}, \begin{pmatrix} 2 & 8 & 4 \\ 8 & 6 & -1 \\ 4 & -1 & 0 \end{pmatrix}.$$

3. For each of the following matrices A, verify that $A^t = -A$:

$$\begin{pmatrix} 0 & 1 & 4 \\ -1 & 0 & 7 \\ -4 & -7 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 2 \\ 1 & 0 & 5 \\ -2 & -5 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 6 & -4 \\ -6 & 0 & 8 \\ 4 & -8 & 0 \end{pmatrix}.$$

4. For each of the following matrices A, verify that $(A^t)^t = A$:

$$\begin{pmatrix} -7 & 6 & 3 \\ 4 & -2 & 0 \end{pmatrix}, \begin{pmatrix} -2 & 7 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -3 & 2 & 8 \\ -9 & 0 & 0 \end{pmatrix}$$

5. If
$$A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 1 & 3 \\ 4 & 1 & 8 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 1 & 0 \\ 2 & -3 & 1 \\ 1 & 1 & -1 \end{pmatrix},$$

verisy that $(A+B)^t = A^t + B^t$, $(AB)^t = B^t A^t$.

6. If
$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}, \text{ verify that}$$

$$AA^t = A^t A = I_2.$$

7. If
$$A = \tfrac{1}{3}\begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ -2 & 2 & -1 \end{pmatrix}, \text{ verify that}$$

$$AA^t = A^t A = I_3.$$

Proof. Let B and C be inverses of a square matrix A. Since B is an inverse of A, therefore,

$$AB = BA = I. \quad \text{...(1)}$$

Again, since C is an inverse of A, therefore,

$$AC = CA = I. \quad \text{..(2)}$$

From (1) we find that

$$C(AB) = CI = C. \quad \text{..(3)}$$

Also, from (2) we find that

$$(CA)B = IB = B. \quad \text{...(4)}$$

Since $\qquad C(AB) = (CA)B,$

therefore, from (3) and (4) it follows that

$$B = C.$$

Because of the above theorem, it is customary as well as proper to talk of *the* inverse of an invertible matrix rather than talking of *an* inverse. The inverse of an invertible matrix A is denoted by A^{-1}.

1·12. DETERMINANTS

In the present section we shall try to obtain a necessary and sufficient condition for the invertibility of a matrix. For this purpose we shall construct a function whose domain is the set M of all square matrices over **R** and whose range is contained in **R**. This function will be called the determinant function (abbreviated as det). For any square matrix A, the value of this function will be called the determinant of A and will be denoted by det A or by $|A|$. If $A=[a_{ij}]$, then det A will be written as $|a_{ij}|$. If

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & & a_{nn} \end{pmatrix},$$

then det A will be denoted by

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

The determinant of an $m \times n$ matrix will be called a determinant of order n. The determinant function will be defined in such a manner that det A=0 iff A is non-invertible, or equivalently A is invertible iff det $A \neq 0$.

1·13. DETERMINANTS OF ORDER ONE

Since a 1×1 matrix (a) is invertible iff $a \neq 0$, therefore, our requirement, namely, A *is invertible iff det* A $\neq 0$, will be satisfied if we have the following definition.

Definition 1·6. *If $A=[a_{11}]$ be a 1×1 matrix, then det $A=a_{11}$.*

1·14. DETERMINANTS OF ORDER TWO

Theorm 1·7. *The 2×2 matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is invertible iff $a_{11}a_{22}-a_{12}a_{21} \neq 0$.*

Proof. $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is invertible

$\Leftrightarrow$ there exists a matrix $\begin{pmatrix} x & y \\ z & w \end{pmatrix}$,

such that

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x & y \\ z & w \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\Leftrightarrow \begin{pmatrix} a_{11}x+a_{12}z & a_{11}y+a_{12}w \\ a_{21}x+a_{22}z & a_{21}y+a_{22}w \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

for some $x, y, z, w \in$ R,

$$\Leftrightarrow \quad \begin{cases} a_{11}x+a_{12}z=1, & a_{11}y+a_{12}w=0, \\ a_{21}x+a_{22}z=0, & a_{21}y+a_{22}w=1, \end{cases}$$

have a common solution,

$$\Leftrightarrow \quad \begin{cases} \triangle x=a_{22}, & \triangle y=-a_{12}, \\ \triangle z=-a_{21}, & \triangle w=a_{11}, \end{cases}$$

have a common solution, where $\triangle=a_{11}a_{22}-a_{12}a_{21}$,

$$\Leftrightarrow \quad a_{11}a_{22}-a_{12}a_{21}\neq 0.$$

Our requirement regarding the determinant function and the above theorem, when taken together, suggest the following definition :

Definition 1·7. *If $A=[a_{ij}]$ be a 2×2 matrix, then*

$$det\ A=a_{11}a_{22}-a_{12}a_{21}.$$

Illustrations.

1. If $\quad A=\begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix},$

then det $A=3 \,.\, 4-2 \,.\, 1=10$.

2. If $\quad A=\begin{pmatrix} a & b \\ c & d \end{pmatrix},$

then det $A=ad-bc$.

Remarks. 1. Observe that (*i*) det A is a sum of 2(=2 !) terms.

(*ii*) Each term in the value $a_{11}a_{22}-a_{12}a_{21}$ contains exactly one element from each row and each column of A.

(*iii*) One term has the sign '+' affixed to it and the other term has the sign '—' affixed to it.

2. We shall very often talk of the determinant

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

instead of using the phrase 'the determinant of the matrix

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

The same will be done later for determinants of higher orders.

1·14·1 Properties of Determinants of Order 2

Property I. *If A be a 2×2 matrix, then*

$$det\ A^t=det\ A.$$

Proof. Let $A=\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix},$

so that

$$A^t=\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}.$$

Now $\det A^t = a_{11}a_{22} - a_{21}a_{12}$,

$$= a_{11}a_{22} - a_{12}a_{21},$$
$$= \det A.$$

The above property is often expressed by saying : *a transposition leaves the value of a determinant unaltered.*

As a consequence of this property, we have the following principle :

Every property which is true of the rows (columns) of a determinant, is also true of its columns (rows).

Property II. *If two rows (columns) of a determinant are proportional (in particular identical), the value of the determinant is zero.*

Proof. Let $a_{21} = ka_{11}$, $a_{22} = ka_{12}$.

Then
$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$
$$= a_{11}a_{22} - a_{12}a_{21},$$
$$= a_{11}(ka_{12}) - a_{12}(ka_{11}),$$
$$= 0.$$

Property III. *If two rows (columns) of a determinant are interchanged, the value of the determinant so obtained is the negative of the value of the original determinant.*

Proof.
$$\begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix} = a_{21}a_{12} - a_{11}a_{22}$$
$$= -\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

Property IV. *If the elements of any row (column) of a determinant are multiplied by the same number k (say), the value of the determinant so obtained is k times the value of the original determinant.*

Proof.
$$\begin{vmatrix} ka_{11} & ka_{12} \\ a_{21} & a_{22} \end{vmatrix} = (ka_{11})a_{22} - (ka_{12})a_{21},$$
$$= k(a_{11}a_{22} - a_{12}a_{21}),$$
$$= k\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

The other case can be considered similarly.

Property V. *If to the elements of a row (column) of a determinant are added k times the corresponding elements of another row (column), the value of the determinant thus obtained is equal to the value of the original determinant.*

Proof.

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21}+ka_{11} & a_{22}+ka_{12} \end{vmatrix} = a_{11}(a_{22}+ka_{12})-a_{12}(a_{21}+ka_{11}),$$
$$= a_{11}a_{22}-a_{12}a_{21},$$
$$= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

The other case can be considered similarly.

EXERCISE 1 (*e*)

1. Find the value of det A for each of the following matrices A :

(*a*) $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, (*b*) $\begin{pmatrix} 4 & -5 \\ 0 & 6 \end{pmatrix}$,

(*c*) $\begin{pmatrix} 7 & -2 \\ 1 & 3 \end{pmatrix}$, (*d*) $\begin{pmatrix} \cos x & \sin x \\ -\sin x & \cos x \end{pmatrix}$.

2. Find the value of each of the following determinants :

(*a*) $\begin{vmatrix} x & y \\ -y & x \end{vmatrix}$ (*b*) $\begin{vmatrix} \sec x & \tan x \\ \tan x & \sec x \end{vmatrix}$.

3. Show that

$$\begin{vmatrix} ad+bc & bd-ac \\ ac-bd & ad+bc \end{vmatrix} = (a^2+b^2)(c^2+d^2).$$

4. Show that

$$\begin{vmatrix} a+ib & c+id \\ -c+id & a-ib \end{vmatrix} = a^2+b^2+c^2+d^2.$$

5. Show that

$$\begin{vmatrix} a\alpha+b\gamma & c\alpha+d\gamma \\ a\beta+b\delta & c\beta+d\delta \end{vmatrix} = (ad-bc)(\alpha\delta-\beta\gamma).$$
$$= \begin{vmatrix} a & b \\ c & d \end{vmatrix} \times \begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix}.$$

6. Show that

$$\begin{vmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}^2.$$

7. If a, b, h are real numbers, show that the roots of the equation

$$\begin{vmatrix} a-x & h \\ h & b-x \end{vmatrix} = 0,$$

are all real.

8. Show that
$$\begin{vmatrix} a_1+c_1 & b_1+d_1 \\ a_2 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} + \begin{vmatrix} c_1 & d_1 \\ a_2 & b_2 \end{vmatrix}.$$

9. Show that
$$\begin{vmatrix} a_1+c_1 & b_1 \\ a_2+c_2 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} + \begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}.$$

Remark. Problem 8 (9) shows that if each element of a row (column) of a determinant is the sum of two numbers, then the determinant can be expressed as the sum of two determinants. This result will be found useful in computations.

1·15. DETERMINANTS OF ORDER THREE

We shall now consider the values of the determinant function for 3×3 matrices. The following theorem will be of help to us in this matter.

Theorem 1·8 *The 3-rowed square matrix* $[a_{ij}]$ *is invertible iff*
$$\triangle = a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}$$
$$-a_{13}a_{22}a_{31}\neq 0.$$

Proof. First, let us assume that the matrix $A=[a_{ij}]$ is invertible and that
$$B=\begin{pmatrix} x & x' & x'' \\ y & y' & y'' \\ z & z' & z'' \end{pmatrix}$$
is the inverse of A. Writing down the identity $AB=I$, we find that x, y, z must satisfy the following equations (and 6 more) :
$$\begin{aligned} a_{11}x+a_{12}y+a_{13}z&=1, \\ a_{21}x+a_{22}y+a_{23}z&=0, \\ a_{31}x+a_{32}y+a_{33}z&=0. \end{aligned} \qquad \text{...(1)}$$

Solving the last two of the above equations, we have
$$\begin{aligned} x&=k(a_{22}a_{33}-a_{23}a_{32}), \\ y&=k(a_{23}a_{31}-a_{21}a_{33}), \\ z&=k(a_{21}a_{32}-a_{22}a_{31}), \end{aligned}$$
where k is any real number. Substituting the values in the first of the equations (1), we have
$$k\triangle=1,$$
and consequently
$$\triangle\neq 0.$$

The condition $\triangle\neq 0$ is sufficient as well. For, if $\triangle\neq 0$, then it can be seen by actual multiplication that the matrix

$$C=\frac{1}{\Delta}\begin{pmatrix} a_{22}a_{33}-a_{23}a_{32} & a_{13}a_{32}-a_{12}a_{33} & a_{12}a_{23}-a_{13}a_{22} \\ a_{23}a_{31}-a_{21}a_{33} & a_{11}a_{33}-a_{13}a_{31} & a_{13}a_{21}-a_{11}a_{23} \\ a_{21}a_{32}-a_{22}a_{31} & a_{12}a_{31}-a_{11}a_{32} & a_{11}a_{22}-a_{12}a_{21} \end{pmatrix}$$

is the inverse of A. (We have only to check up that AC=CA=I).

In view of the above theorem we can have the following definition :

Definition 1·8. *If* $A=[a_{ij}]$ *be a* 3×3 *matrix, then*

$$det\ A=a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}-a_{13}a_{22}a_{31}.$$

The following points may be noted in respect of the above definition :

(*i*) det A is the sum of 6(=3 !) terms out of which 3(=3 !)/2) have the sign '+' affixed to them and the other 3 terms have the sign '−' affixed to them.

(*ii*) Each term in det A is the product of three elements of A, exactly one of which belongs to each row and exactly one of which belongs to each column.

(*iii*) It can be easily verified that $A=\Sigma\pm a_{1i}a_{2j}a_{3k}$ where on the right hand side we have to take all the terms obtained by giving to *i, j, k* all possible distinct values from among 1, 2, 3 and the sign of the term $a_{1i}a_{2j}a_{3k}$ is + or − according as *i, j, k* is a cyclic or an anti-cyclic arrangement of 1, 2, 3.

1·15·1. Sarrus Diagram

We shall now give a diagram, due to Sarrus, which is useful in remembering the expression for the value of a determinant of order 3.

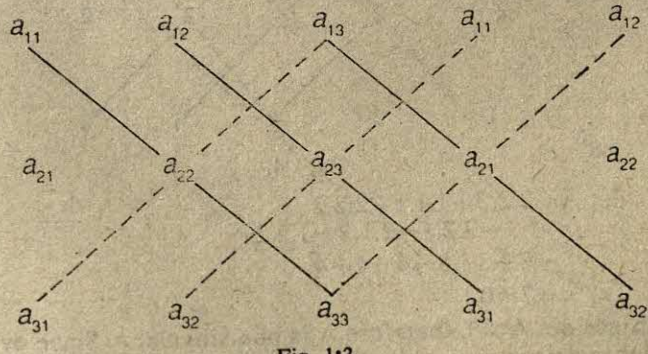


Fig. 1·3.

The terms prefixed with the '+' sign in definition 1-8 are those which correspond to the elements joined by continuous lines and the terms prefixed with the '−' sign in the same definition are those which correspond to the elements joined by dotted lines in the above diagram

1·16. SINGULAR AND NON-SINGULAR MATRICES

Definition 1·9. *A square matrix A is said to be non-singular if* $|A| \neq 0$ *; it is said to be singular if* $|A| = 0$.

In view of the above definition, theorem 1·8 can be restated as follows :

The three-rowed square matrix A is invertible iff it is non-singular.

The above theorem is true for square matrices of all orders. The proof in the general case is, however, beyond the scope of the present book.

From the above discussion we find that to test whether a given matrix is non-singular (or invertible) we have simply to evaluate the determinant of the matrix.

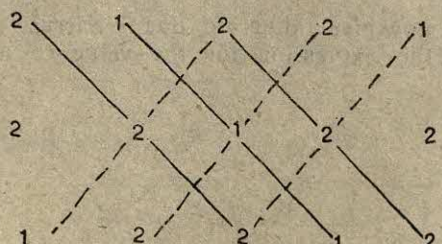
Example 6. *Show that the matrix*

$$\begin{pmatrix} 2 & 1 & 2 \\ 2 & 2 & 1 \\ 1 & 2 & 2 \end{pmatrix}$$

is non-singular. Is it invertible ?

Solution. Let us denote the given matrix by A. We shall compute det A.

Sarrus diagram for A is



Fig, 1·4.

$$\begin{aligned} \det A &= 2.2.2 + 1.1.1 + 2.2.2 \\ &\quad - 2.2.1 - 2.1.2 - 1.2.2 \\ &= 8 + 3 + 8 - 4 - 4 - 4 \\ &= 7 \neq 0. \end{aligned}$$

Since det $A \neq 0$, therefore A is non-singular. Since every non-singular matrix is invertible, therefore A is invertible.

EXERCISE 1 (*f*)

Evaluate each of the following determinants :

1. $\begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{vmatrix}$ 2. $\begin{vmatrix} 2 & -1 & 4 \\ 3 & 2 & 20 \\ -1 & 2 & 4 \end{vmatrix}$.

3. $\begin{vmatrix} -1 & 1 & 0 \\ 0 & 2 & 3 \\ 3 & 1 & 0 \end{vmatrix}$ 4. $\begin{vmatrix} 1 & 1 & 1 \\ -1 & -2 & -3 \\ 1 & 4 & 9 \end{vmatrix}$.

Show that each of the following matrices is non-singular :

5. $\begin{pmatrix} 1 & -2 & 3 \\ 0 & -1 & 4 \\ -2 & 2 & 1 \end{pmatrix}$ 6. $\begin{pmatrix} 1 & 2 & -1 \\ -4 & -7 & 4 \\ -4 & -9 & 5 \end{pmatrix}$.

Show that each of the following matrices is invertible :

7. $\begin{pmatrix} 3 & -15 & 5 \\ -1 & 6 & -2 \\ 1 & -5 & 2 \end{pmatrix}$ 8. $\begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 3 \\ 4 & 6 & 2 \end{pmatrix}$.

Examine whether the following matrices are singular or non-singular :

9. $\begin{pmatrix} 1 & 3 & -2 \\ 5 & 3 & 7 \\ 7 & 9 & 3 \end{pmatrix}$ 10. $\begin{pmatrix} 1 & 1 & 2 \\ 3 & 1 & -3 \\ 2 & 1 & -1 \end{pmatrix}$.

1·17. PROPERTIES OF DETERMINANTS OF ORDER 3

Property I. *If A be a 3×3 matrix, then det A^t=det A.*

Proof. Let $A=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$

Then $A^t=\begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$.

The Sarrus diagram for A^t is

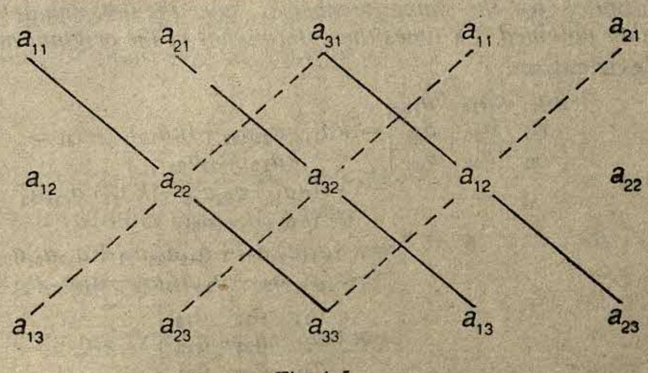


Fig. 1·5.

From the above diagram, we find that

$$\det A^t = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33},$$
$$= \det A.$$

The above property is often expressed by saying that *transposition leaves the value of a determinant unaltered.*

As a consequence of the above property, we have the following principle :

Every theorem which is true of the rows (columns) of a determinant is also true of its columns (rows).

Property II. *If two rows (columns) of a determinant are proportional (in particular identical), the value of the determinant is zero.*

Verification. Let $a_{31}=ka_{11}$, $a_{32}=ka_{12}$, $a_{33}=ka_{13}$.

Then
$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31},$$
$$= k\,(a_{11}a_{22}a_{13} + a_{12}a_{23}a_{11} + a_{13}a_{21}a_{12} - a_{11}a_{23}a_{12} - a_{12}a_{21}a_{13} - a_{13}a_{22}a_{11}),$$
$$=0.$$

Property III. *If two rows (columns) of a determinant are interchanged, the value of the determinant so obtained is the negative of the value of the original determinant.*

Verification.

$$\begin{vmatrix} a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \end{vmatrix} = a_{31}a_{22}a_{13} + a_{32}a_{23}a_{11} + a_{33}a_{21}a_{12} - a_{31}a_{23}a_{12} - a_{32}a_{21}a_{13} - a_{33}a_{22}a_{11},$$
$$=(a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}),$$
$$= - \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Property IV. *If the elements of a row (column) of a matrix are multiplied by the same number k, say, the determinant of the matrix thus obtained is k times the determinant of the original matrix.*

Verification.

$$\begin{vmatrix} ka_{11} & ka_{12} & ka_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = (ka_{11})\,a_{22}a_{33} + (ka_{12})\,a_{23}a_{31} + (ka_{13})\,a_{21}a_{32} - (ka_{11})\,a_{23}a_{32} - (ka_{12})\,a_{21}a_{33} - (ka_{13})\,a_{22}a_{31},$$
$$= k\,(a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{31}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}),$$
$$= k \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Property V. *If to the elements of a row (column) of a determinant are added k times the corresponding elements of another row (column), the value of the determinant thus obtained is equal to the value of the original determinant.*

Verification. We shall verify the result only for one case, namely when k times the elements of the second row are added to the corresponding elements of the third row.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31}+ka_{21} & a_{32}+ka_{22} & a_{33}+ka_{23} \end{vmatrix}$$

$$=a_{11}a_{22}(a_{33}+ka_{23})+a_{12}a_{23}(a_{31}+ka_{21})+a_{13}a_{21}(a_{32}+ka_{22})$$
$$-a_{11}a_{23}(a_{32}+ka_{22})-a_{12}a_{21}(a_{33}+ka_{23})-a_{13}a_{22}(a_{31}+ka_{21}),$$

$$=a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}-a_{13}a_{22}a_{31},$$

$$=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

EXERCISE 1 (*q*)

1. Explain (without evaluating) why the following two determinants are equal. Verify this fact by actual evaluation.

$$\begin{vmatrix} 3 & 7 & 1 \\ -2 & 1 & 4 \\ 6 & -4 & 3 \end{vmatrix} \text{ and } \begin{vmatrix} 3 & -2 & 6 \\ 7 & 1 & -4 \\ 1 & 4 & 3 \end{vmatrix}$$

Without evaluating, state why each of the determinants in problems 2 and 3 is zero. Check by evaluating.

2. $\begin{vmatrix} 1 & 2 & 4 \\ -3 & 1 & 2 \\ 1 & 2 & 4 \end{vmatrix}.$ 3. $\begin{vmatrix} 2 & 8 & 4 \\ -5 & 6 & -10 \\ 1 & 7 & 2 \end{vmatrix}.$

4. Show by actual expansion that the sign of the determinant

$$\begin{vmatrix} -1 & 5 & 6 \\ 2 & 3 & 1 \\ 1 & 0 & 4 \end{vmatrix}$$

is changed if the second and third rows are interchanged.

5. State (without evaluating) why

$$\begin{vmatrix} 2 & 1 & 3 \\ -3 & 1 & 4 \\ 2 & 2 & 1 \end{vmatrix} = -\begin{vmatrix} 3 & 1 & 2 \\ 4 & 1 & -3 \\ 1 & 2 & 2 \end{vmatrix}.$$

Verify this fact by actual evaluation.

6. State (without evaluating) why

$$\begin{vmatrix} 3 & 7 & 4 \\ -2 & 1 & 5 \\ 6 & 18 & 3 \end{vmatrix} = 3\begin{vmatrix} 3 & 7 & 4 \\ -2 & 1 & 5 \\ 2 & 6 & 1 \end{vmatrix}.$$

Verify this fact by actual evaluation.

7. In the following determinant, multiply each of the elements in the second row by 4, and form a new determinant by adding these products to the corresponding elements of the third row. Show by actual evaluation that the value of the new determinant is equal to that of the original determinant :

$$\begin{vmatrix} 3 & -1 & 2 \\ -2 & 3 & 4 \\ 1 & 5 & -3 \end{vmatrix}.$$

8. Show that

$$\begin{vmatrix} a_1+\alpha_1 & b_1 & c_1 \\ a_2+\alpha_2 & b_2 & c_2 \\ a_3+\alpha_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} \alpha_1 & b_1 & c_1 \\ \alpha_2 & b_2 & c_2 \\ \alpha_3 & b_3 & c_3 \end{vmatrix}.$$

9. Show that

$$\begin{vmatrix} a_1+\alpha_1 & b_1+\beta_1 & c_1+\gamma_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_2 & b_3 & c_3 \end{vmatrix} + \begin{vmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

Remark. Problems 8 and 9 show that if each element of a column or row of a determinant is the sum of two numbers, then the determinant can be expressed as the sum of two determinants. This result will be found useful in computations.

MINORS

Consider the matrix

$$A=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Let us recall that if we strike off any one row and any one column of A, then the 2×2 matrix thus obtained is a sub-matrix of A. Thus, for example, if we strike off the third row and the second column, we get the sub-matrix

$$\begin{pmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{pmatrix}$$

The determinant of any such sub-matrix is called a minor of det A. Thus

$$\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \text{ is a minor of det A.}$$

In the above manner we can get 3^2 minors. The minor obtained by striking off the pth row and the qth column of a matrix $A=[a_{ij}]$ is called the minor of a_{pq} in det A. Thus

$$\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \text{ is the minor of } a_{32}.$$

The minors of a_{11}, a_{12}, a_{13} in $|a_{ij}|$ are

$$\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}, \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

respectively.

The minors of a_{21}, a_{22}, a_{23} in $|a_{ij}|$ are

$$\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}, \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} \text{ and } \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}.$$

respectively.

The minors of a_{31}, a_{32}, a_{33} in $|a_{ij}|$ are

$$\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}, \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \text{ and } \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

respectively.

Theorem 1·9. *A determinant of order 3 can be expressed as a linear combination of the minors of the elements of any row or column.*

Verification. We shall show that

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

is expressible as a linear combination of the elements of the first row.

Denoting the given determinant by det A, we have

$$\det A = a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}-a_{13}a_{22}a_{31},$$

$$=a_{11}(a_{22}a_{33}-a_{23}a_{32})-a_{12}(a_{21}a_{33}-a_{23}a_{31})+a_{13}(a_{21}a_{32}-a_{22}a_{31}),$$

$$=a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}-a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}+a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$=a_{11}\text{ (minor of } a_{11})-a_{12}\text{ (minor of } a_{12})+a_{13}\text{ (minor of } a_{13}).$$

The above expression for det A is usually called the *expansion of det A in terms of the minors of elements of the first row.*

Similarly,

$$\det A=-a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}+a_{22}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}-a_{23}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

$$=-a_{21}\text{ (minor of } a_{21})+a_{22}\text{ (minor of } a_{22})-a_{23}\text{ (minor of } a_{33}).$$

Also,

$$\det A=a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}-a_{32}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}+a_{33}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

$$=a_{31}\text{ (minor of } a_{31})-a_{32}\text{ (minor of } a_{32})+a_{33}\text{ (minor of } a_{33})$$

The expressions for det A in terms of the minors of the elements of the first, second and third columns are

$$a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix},$$

$$-a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{22}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{32}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix},$$

$$a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} - a_{23}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} + a_{33}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},$$

respectively.

To evaluate a determinant, we often expand it in terms of the minors of the elements of a row or column. If a row or column contains one or more zeros (as elements), it is convenient to expand the determinant in terms of the elements of that row or column. Before expanding it is useful to make use of the properties I—V studied already in such a manner that some of the elements become zero (or small in absolute value), The following examples will illustrate the technique.

Example 7. *Evaluate the determinant*

$$\begin{vmatrix} 1 & -1 & 2 \\ 3 & 4 & 7 \\ 2 & 8 & 5 \end{vmatrix}$$

by expanding it in terms of the minors of the elements of the first row.

Solution.

$$\begin{vmatrix} 1 & -1 & 2 \\ 3 & 4 & 7 \\ 2 & 8 & 5 \end{vmatrix}$$

$$=1\begin{vmatrix} 4 & 7 \\ 8 & 5 \end{vmatrix} - (-1)\begin{vmatrix} 3 & 7 \\ 2 & 5 \end{vmatrix} + 2\begin{vmatrix} 3 & 4 \\ 2 & 8 \end{vmatrix}$$

$$=1\ (4.5-7.8)-(-1)(3.5-7.2)+2\ (3.8-4.2)$$

$$=-36+1+32$$

$$=-3.$$

Example 8. *Evaluate the determinant*

$$\begin{vmatrix} 20 & 42 & 89 \\ 44 & 93 & 198 \\ 93 & 191 & 450 \end{vmatrix}$$

Solution.

$$\begin{vmatrix} 20 & 42 & 89 \\ 44 & 93 & 198 \\ 93 & 191 & 450 \end{vmatrix}$$

$$=\begin{vmatrix} 20 & 42 & 89 \\ 4 & 9 & 20 \\ 13 & 23 & 94 \end{vmatrix}$$ [Subtracting 2 times the 1st row from the 2nd row, and subtracting 4 times the 1st row from the 3rd row]

$$=\begin{vmatrix} 20 & 2 & 9 \\ 4 & 1 & 4 \\ 13 & -3 & 42 \end{vmatrix}$$ [Subtracting 2 times the first column from the second column, and 4 times the first column from the third column]

$$=\begin{vmatrix} 12 & 0 & 1 \\ 4 & 1 & 4 \\ 25 & 0 & 54 \end{vmatrix}$$ [Subtracting 2 times the second row from the first row and adding 3 times the second row from the third row]

$$=\begin{vmatrix} 12 & 1 \\ 25 & 54 \end{vmatrix}$$ [Expanding in terms of the minors of the elements of the second column]

$=12.54-1.25$

$=623.$

1·18. COFACTORS

If we multiply the minor of the elements in the pth row and the qth column of the determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

by $(-1)^{p+q}$, the product is called the co-factor of that element. It is usual to denote the co-factor of an element by the corresponding capital letter. With this notation,

$$A_{11}=(-1)^{1+1}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \quad A_{12}=(-1)^{1+2}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$$

$$A_{13}=(-1)^{1+3}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}, \quad A_{21}=(-1)^{2+1}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}$$

$$A_{22}=(-1)^{2+2}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \quad A_{23}=(-1)^{2+3}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

$$A_{31}=(-1)^{3+1}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}, \quad A_{32}=(-1)^{3+2}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}$$

$$A_{33}=(-1)^{3+2}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

From the above expressions for co-factors we find that the co-factor of an element differs from its minor by a factor -1 at the most.

Theorem 1·10. *If $A=[a_{ij}]$ be a 3×3 matrix, then*

(*i*) $$\sum_j a_{ij}A_{pj}=\begin{cases} \det A, & \text{if } i=p, \\ 0, & \text{if } i\neq p. \end{cases}$$

(*ii*) $\sum_{i} a_{ij} A_{iq} = \begin{cases} \det A, & \text{if } j=q. \\ 0, & \text{if } j \neq q. \end{cases}$

Proof.

(*i*) $\sum_{j} a_{1j} A_{1j} = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}$,

$$= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix},$$

$=\det A.$

Similarly we can show that

$$\sum_{j} a_{2j} A_{2j} = \det A,$$

$$\sum_{j} a_{3j} A_{3j} = \det A.$$

Again, $\sum_{j} a_{1j} A_{2j} = a_{11}A_{21} + a_{12}A_{22} + a_{13}A_{23}$,

$$= -a_{11}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{13}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

$$= -\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$=0$, since the first and second rows are identical.

Similarly we can show that

$$\sum_{i} a_{1j}A_{3j} = \sum_{j} a_{2j}A_{1j} = \sum_{j} a_{2j}A_{3j} = \sum_{j} a_{3j}A_{1j}$$

$$= \sum_{j} a_{3j}A_{2j} = 0.$$

(*ii*) The proof is left as an exercise for the reader.

Remark. The contents of the above theorem are usually expressed by saying that

(*i*) *the sum of the products of the elements of any row or column of a determinant with the corresponding cofactors is equal to the value of the determinant* ;

(*ii*) *the sum of the products of the elements of a row* (*column*) *with the co-factors of the corresponding elements of any other row* (*column*) *is zero.*

1·19. SOME ASSORTED EXAMPLES ON DETERMINANTS

Here below we give some assorted examples which illustrate the method of evaluating determinants of the third order.

Example 9. *Show that*

$$=\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix}=(b-c)(c-a)(a-b).$$

Solution. $\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix}$

$$=\begin{vmatrix} 0 & 0 & 1 \\ a-c & b-c & c \\ a^2-c^2 & b^2-c^2 & c^2 \end{vmatrix}$$ [Subtracting the third column from the first and the second column].

$$=(a-c)(b-c)\begin{vmatrix} 0 & 0 & 1 \\ 1 & 1 & c \\ a+c & b+c & c^2 \end{vmatrix}$$ [Taking $a-c$ common from the first column and $b-c$ common from the second column].

$$=(a-c)(b-c)\begin{vmatrix} 1 & 1 \\ a+c & b+c \end{vmatrix}$$ [Expanding in terms of the elements of the first row].

$$=(a-c)(b-c)\{(b+c)-(a+c)\}$$
$$=(a-c)(b-c)(b-a)$$
$$= b-c)(c-a)(a-b)$$

Alternative Solution. Let

$$\triangle=\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix}.$$

Regarding $\triangle$ as a polynomial in a, we find that $\triangle$ vanishes when $a=b$ (for then the first and the second columns become identical). Therefore, $a-b$ must be a factor of $\triangle$.

Similarly $b-c$, $c-a$ are also factors of $\triangle$.

Since each term in the expression for $\triangle$ is of degree 3 in a, b, c, therefore, we must have

$$\triangle=\lambda(b-c)(c-a)(a-b),$$

where λ is a constant. To find the value of λ, we compare the coefficients of bc^2 on both sides. This yields $\lambda=1$.

Hence $\triangle=(b-c)(c-a)(a-b)$.

Example 10. *Show that*

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^3 & b^3 & c^3 \end{vmatrix} = (b-c)(c-a)(a-b)(a+b+c).$$

Solution.

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^3 & b^3 & c^3 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 0 & 0 \\ a & b-a & c-a \\ a^3 & b^3-a^3 & c^3-a^3 \end{vmatrix}$$ [Subtracting the first column from the second and the third column].

$$= \begin{vmatrix} b-a & c-a \\ b^3-a^3 & c^3-a^3 \end{vmatrix}$$ [Expanding in terms of the elements of the first row].

$$=(b-a)(c-a) \begin{vmatrix} 1 & 1 \\ b^2+ab+a^2 & c^2+ac+a^2 \end{vmatrix}$$

[Taking $b-a$ common from the first column and $c-a$ common from the second column],

$=(b-a)(c-a)\{(c^2+ac+a^2)-(b^2+ab+a^2)\}$

$=(b-a)(c-a)\{(c^2-b^2)+a(c-b)\}$,

$=(b-a)(c-a)(c-b)\{(c+b+a)\}$,

$=(b-c)(c-a)(a-b)(a+b+c)$.

Alternative Solution. Let the given determinant be denoted by Δ. Regarding Δ as a polynomial in a, we find that Δ vanishes when $a=b$ (for then the first and the second columns become identical). Therefore, $a-b$ must be a factor of Δ. Similarly $b-c$ and $c-a$ must also be factors of Δ. Since each term in the expression for Δ is of degree 4 in a, b, c, we must have

$$\Delta \equiv f(a, b, c)(b-c)(c-a)(a-b),$$

where $f(a, b, c)$ is a homogeneous expression of the first degree in a, b, c.

Also, since Δ and $(b-c)(c-a)(a-b)$ both remain unaltered by a cyclic interchange of a, b, c (*i.e.* when we replace a, b, c by b, c and a respectively), therefore $f(a, b, c)$ must also remain unaltered by such an interchange. Therefore, $f(a, b, c)$ must be of the form $k(a+b+c)$, where k is a constant.

Therefore, $\Delta = k(b-c)(c-a)(a-b)(a+b+c)$ ...(*i*)

Since (*i*) must hold for all values of a, b, c therefore, by putting $a=0$, $b=1$, $c=2$ on both sides of (*i*), we have

$$\begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 8 \end{vmatrix} = k(-1).2.(-1).3,$$

or $$6=6k,$$
or $$k=1.$$
Hence $$\Delta=(b-c)(c-a)(a-b)(a+b+c).$$

Remark. We could have also determined k by comparing the coefficient of bc^3 on both sides of (i).

Example 11. *Prove that*

$$\begin{vmatrix} 1 & a^2+bc & a^3 \\ 1 & b^2+ca & b^3 \\ 1 & c^2+ab & c^3 \end{vmatrix} = -(b-c)(c-a)(a-b)(a^2+b^2+c^2).$$

Solution. Denoting the given determinant by Δ, we can immediately show, as in Examples 9 and 10 that $b-c$, $c-a$, $a-b$ must be factors of Δ. Since each term in the expression for Δ is of the fifth degree in a, b, c, therefore,

$$\Delta=f(a, b, c)(b-c)(c-a)(a-b),$$

where $f(a, b, c)$ is a homogeneous expression of the second degree in a, b, c. Since both Δ and $(b-c)(c-a)(a-b)$ are unaffected by a cyclic interchange of a, b, c, therefore, $f(a, b, c)$ must also remain unaffected by such an interchange.

$$\therefore \quad f(a, b, c)=(a^2+b^2+c^2)+\mu(bc+ca+ab),$$

where λ and μ are constants. Therefore,

$$\Delta=(b-c)(c-a)(a-b)\,\{\lambda(a^2+b^2+c^2)+\mu(bc+ca+ab)\} \qquad \ldots(i)$$

Since (i) holds for all values of a, b, c, therefore, on putting $a=0$, $b=1$, $c=2$, we have

$$\begin{vmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 4 & 8 \end{vmatrix} = (-1).2(-1)(5\lambda+2\mu),$$

or $$5\lambda+2\mu=-5. \qquad \ldots(ii)$$

Similarly by putting $a=0$, $b=1$, $c=-2$, we have

$$5\lambda-2\mu=-5. \qquad \ldots(iii)$$

Solving (ii) and (iii) for λ and μ, we have

$$\lambda=-1,\ \mu=0.$$

Hence $$\Delta=-(b-c)(c-a)(a-b)(a^2+b^2+c^2).$$

Example 12. *Show that*

$$\begin{vmatrix} 1 & \beta\gamma+\alpha\delta & \beta^2\gamma^2+\alpha^2\delta^2 \\ 1 & \gamma\alpha+\beta\delta & \gamma^2\alpha^2+\beta^2\delta^2 \\ 1 & \alpha\beta+\gamma\delta & \alpha^2\beta^2+\gamma^2\delta^2 \end{vmatrix} = -(\beta-\gamma)(\gamma-\alpha)(\alpha-\beta)(\alpha-\delta)(\beta-\delta)(\gamma-\delta)$$

Solution. Adding $2\alpha\beta\gamma\delta$ times the elements of the first column to the crr esponding elements of the third column, the given determinant equals

$$\begin{vmatrix} 1 & \beta\gamma+\alpha\delta & (\beta\gamma+\alpha\delta)^2 \\ 1 & \gamma\alpha+\beta\delta & (\gamma\alpha+\beta\delta)^2 \\ 1 & \alpha\beta+\gamma\delta & (\alpha\beta+\gamma\delta)^2 \end{vmatrix}.$$

If we put $\beta\gamma+\alpha\delta=a$, $\gamma\alpha+\beta\delta=b$, $\alpha\beta+\gamma\delta=c$, the determinant equals

$$\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix}=(b-c)(c-a)(a-b),$$

by Example 9.

Since $b-c=(\gamma\alpha+\beta\delta)-(\alpha\beta+\gamma\delta)=-(\beta-\gamma)(\alpha-\delta)$,

$c-a=(\alpha\beta+\gamma\delta)-(\beta\gamma+\alpha\delta)=-(\gamma-\alpha)(\beta-\delta)$,

$a-b=(\beta\gamma+\alpha\delta)-(\gamma\alpha+\beta\delta)=-(\alpha-\beta)(\gamma-\delta)$,

therefore, the given determinant

$$=(b-c)(c-a)(a-b),$$
$$=-(\beta-\gamma)(\gamma-\alpha)(\alpha-\beta)(\alpha-\delta)(\beta-\delta)(\gamma-\delta)$$

Example 13. *Show that*

$$\begin{vmatrix} (b+c)^2 & a^2 & a^2 \\ b^2 & (c+a)^2 & b^2 \\ c^2 & c^2 & (a+b)^2 \end{vmatrix}=2abc(a+b+c)^3.$$

Solution.

$$=\begin{vmatrix} (b+c)^2 & a^2 & a^2 \\ b^2 & (c+a)^2 & b^2 \\ c^2 & c^2 & (a+b)^2 \end{vmatrix}$$

$$\begin{vmatrix} (b+c)^2-a^2 & 0 & a^2 \\ 0 & (c+a)^2-b^2 & b^2 \\ c^2-(a+b)^2 & c^2-(a+b)^2 & (a+b)^2 \end{vmatrix}$$

[Subtracting the third column from the first and the second column],

$$=(a+b+c)^2\begin{vmatrix} b+c-a & 0 & a^2 \\ 0 & c+a-b & b^2 \\ c-a-b & c-a-b & (a+b)^2 \end{vmatrix}$$

[Taking $a+b+c$ common from each of the columns (*i*) and (*ii*)],

$$=(a+b+c)^2\begin{vmatrix} b+c-a & 0 & a^2 \\ 0 & c+a-b & b^2 \\ -2b & -2a & 2ab \end{vmatrix}$$

[Subtracting rows (*i*) and (*ii*) from row (*iii*)],

$$=\frac{(a+b+c)^2}{ab}\begin{vmatrix} (b+c)a & a^2 & a^2 \\ b^2 & (c+a)b & b^2 \\ 0 & 0 & 2ab \end{vmatrix}$$

[Adding col. (*iii*) to a times col. (*i*) and to b times col. (*ii*)],

$$=\frac{(a+b+c)^2}{ab}.2ab\begin{vmatrix} (b+c)a & a^2 \\ b^2 & (c+a)b \end{vmatrix}$$

[Expanding in terms of the elements of the third row],

$$=2ab\,(a+b+c)^2 \begin{vmatrix} b+c & a \\ b & c+a \end{vmatrix}$$

[Taking a common from the first row and b common from the second row],

$$=2ab(a+b+c)^2\,\{(b+c)(c+a)-ab\},$$
$$=2abc(a+b+c)^3.$$

EXERCISE 1 (*h*)

Evaluate each of the following determinants :

1. $\begin{vmatrix} 15 & 11 & 9 \\ 12 & 14 & 17 \\ 18 & 13 & 12 \end{vmatrix}$. 2. $\begin{vmatrix} 11 & 118 & 135 \\ 9 & 97 & 113 \\ 4 & 45 & 50 \end{vmatrix}$.

3. $\begin{vmatrix} 60 & 70 & 70 \\ 74 & 71 & 73 \\ 75 & 68 & 72 \end{vmatrix}$. 4. $\begin{vmatrix} 18 & 40 & 107 \\ 40 & 89 & 238 \\ 89 & 198 & 529 \end{vmatrix}$.

5. Prove that

$$\begin{vmatrix} 1 & 1 & 1 \\ \alpha & \beta & \gamma \\ \beta\gamma & \gamma\alpha & \alpha\beta \end{vmatrix}=(\beta-\gamma)(\gamma-\alpha)(\alpha-\beta).$$

6. Prove that

$$\begin{vmatrix} 1 & \beta+\gamma & \beta^2+\gamma^2 \\ 1 & \gamma+\alpha & \gamma^2+\alpha^2 \\ 1 & \alpha+\beta & \alpha^2+\beta^2 \end{vmatrix}=(\beta-\gamma)(\gamma-\alpha)(\alpha-\beta).$$

7. Prove that

$$\begin{vmatrix} \alpha & \beta & \gamma \\ \alpha^2 & \beta^2 & \gamma^2 \\ \beta+\gamma & \gamma+\alpha & \alpha+\beta \end{vmatrix}=(\beta-\gamma)(\gamma-\alpha)(\alpha-\beta)(\alpha+\beta+\gamma).$$

8. Prove that

$$\begin{vmatrix} 1 & 1 & 1 \\ \alpha^2 & \beta^2 & \gamma^2 \\ \alpha^3 & \beta^3 & \gamma^3 \end{vmatrix}=(\beta-\gamma)(\gamma-\alpha)(\alpha-\beta)(\beta\gamma+\gamma\alpha+\alpha\beta).$$

9. Prove that

$$\begin{vmatrix} a-b-c & 2a & 2a \\ 2b & b-c-a & 2b \\ 2c & 2c & c-a-b \end{vmatrix}=(a+b+c)^3.$$

10. Prove that

$$\begin{vmatrix} a^2 & bc & ac+c^2 \\ a^2+ab & b^2 & ac \\ ab & b^2+bc & c^2 \end{vmatrix}=4a^2b^2c^2.$$

11. Prove that

$$\begin{vmatrix} -2a & a+b & a+c \\ b+a & -2b & b+c \\ c+a & c+b & -2c \end{vmatrix}=4(b+c)(c+a)(a+b).$$

12. If a, b, c are all different and if

$$\begin{vmatrix} a & a^2 & a^3+1 \\ b & b^2 & b^3+1 \\ c & c^2 & c^3+1 \end{vmatrix}=0,$$

prove that $abc+1=0$.

13. Prove that

$$\begin{vmatrix} 1+a^2-b^2 & 2ab & -2b \\ 2ab & 1-a^2+b^2 & 2a \\ 2b & -2a & 1-a^2-b^2 \end{vmatrix}=(1+a^2+b^2)^3.$$

14. Prove that

$$\begin{vmatrix} a^2 & a^2-(b-c)^2 & bc \\ b^2 & b^2-(c-a)^2 & ca \\ c^2 & c^2-(a-b)^2 & ab \end{vmatrix}=(b-c)(c-a)(a-b)(a+b+c)(a^2+b^2+c^2).$$

15. Prove that

$$\begin{vmatrix} x+a & b & c \\ a & x+b & c \\ a & b & x+c \end{vmatrix}=x^2\,(x+a+b+c).$$

16. Solve the equation :

$$\begin{vmatrix} 3x-8 & 3 & 3 \\ 3 & 3x-8 & 3 \\ 3 & 3 & 3x-8 \end{vmatrix}=0.$$

17. Find the value of the determinant

$$\begin{vmatrix} 1/a & a^2 & bc \\ 1/b & b^2 & ca \\ 1/c & c^2 & ab \end{vmatrix}$$

18. Prove that

$$\begin{vmatrix} bc & a & a^2 \\ ca & b & b^2 \\ ab & c & c^2 \end{vmatrix}=\begin{vmatrix} 1 & a^2 & a^3 \\ 1 & b^2 & b^3 \\ 1 & c^2 & c^3 \end{vmatrix}.$$

1·20. APPLICATION OF DETERMINANTS TO AREA OF A TRIANGLE

The formula for finding the area of a triangle, the co ordinates of whose vertices are given, can be put in a very neat (and easy to remember) form in terms of a determinant.

Let $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ be the co-ordinates of the vertices A, B and C respectively of a triangle ABC. If $\triangle$ denotes the area of the triangle ABC, then we know that

$$\begin{aligned}\triangle &= \tfrac{1}{2}\{x_1y_2-x_2y_1+x_2y_3-x_3y_2+x_3y_1-x_1y_3\} \\ &= \tfrac{1}{2}\{x_1y_2-x_1y_3-x_2y_1+x_3y_1+x_2y_3-x_3y_2\} \\ &= \tfrac{1}{2}\{x_1(y_2-y_3)-y_1(x_2-x_3)+x_2y_3-x_3y_2\} \qquad ...(i)\end{aligned}$$

$$=\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}.$$

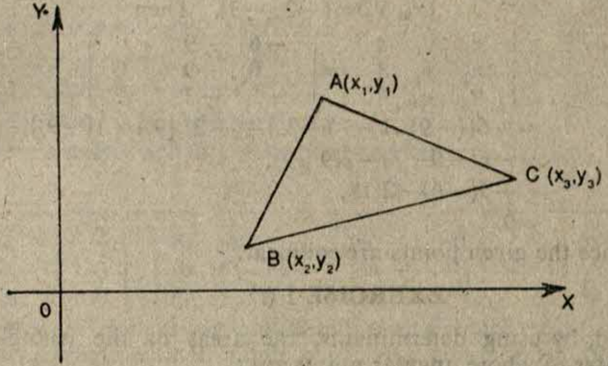


Fig. 1 6.

Corollary. If the points A (x_1, y_1), B (x_2, y_2) and C (x_3, y_3) are collinear, then from (*i*), we find that we must have

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}=0, \qquad ...(ii)$$

which is the required condition.

Example 14. *Find, using determinants, the area of the triangle, the co-ordinates of whose angular points are*

(2, 6), (−4, −2) and (3, −1).

Solution. Let $(x_1, y_1)=(2, 6)$, $(x_2, y_2)=(-4, -2)$,

$(x_3, y_3)=(3, -1)$.

The required area

$$=\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix},$$

$$=\frac{1}{2}\begin{vmatrix} 2 & 6 & 1 \\ -4 & -2 & 1 \\ 3 & -1 & 1 \end{vmatrix},$$

$$=\frac{1}{2}\{2(-2+1)-6(-4-3)+1 \,.\, (4+6)\},$$

$$=\frac{1}{2}\{-2+42+10\},$$

$$=25.$$

Example 15. *Show that the points*

$(-6, 9)$ $(0, -9)$ and $(-2, -3)$ are collinear.

Solution. Take $(x_1, y_1)=(-6, 9)$, $(x_2, y_2)=(0, -9)$ and $(x_3, y_3)=(-2, -3)$. Then

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = \begin{vmatrix} -6 & 9 & 1 \\ 0 & -9 & 1 \\ -2 & -3 & 1 \end{vmatrix}$$

$$=-6\{(-9).1-1(-3)\}+(-2)\{9.1-1(-9)\},$$

$$=-6(-9+3)-2(9+9),$$

$$=-6(-6)-2.18,$$

$$=0.$$

Hence the given points are collinear.

EXERCISE 1 (*i*)

Find, by using determinants, the areas of the triangles the co-ordinates of whose angular points are :

1. $(2, 1), (4, 4), (6, -2)$.
2. $(-1, 2), (1, 3), (2, -4)$.
3. $(9, 3), (-5, -2), (3, 5)$.
4. $(-2, 5), (7, -4), (3, 2)$.
5. $(5, 7), (-2, -1), (0, 8)$.
6. $(-2, -5), (5, 3), (3, 9)$.
7. $(c, a), (c-a, a), (c+a, a)$.
8. $(ap^2, 2ap), (aq^2, 2aq), ar^2, 2ar)$.
9. $(a \cos \alpha, a \sin \alpha), (a \cos \beta, a \sin \beta), (a \cos \gamma, a \sin \gamma)$.
10. $(cp, a/p), (cq, c/q), (cr, c/r)$.

Prove, by using determinants, that the following sets of points are in a straight line :

11. $(2, 3), (5, 7), (8, 11)$.
12. $(5, 1), (11, 4), (1, -1)$.
13. $(2, -3), (-6, 9), (-2, 3)$.
14. $(8, -1). (4, 7), (6, 3)$.
15. $(3, 1), (7, -3), (5, -1)$.

1·21. APPLICATION OF DETERMINANTS TO SOLUTION OF EQUATIONS (*CRAMER'S RULE*)

The following theorem, commonly known as Cramer's rule, tells us as to how we can use determinants to solve a system of three linear equations in three variables. Even though we have stated and proved the theorem for the case of three variables, the theorem is perfectly general and can be extended to a system of n linear equations in n variables.

Theorem 1·11. (Cramer's Rule)

$$\textit{If } \Delta=\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \neq 0,$$

then the system of linear equations

$$a_1+b_1y+c_1z+d_1=0,$$
$$a_2x+b_2y+c_2z+d_2=0,$$
$$a_3x+b_3y+c_3z+d_3=0,$$

possesses a unique solution, which is given by

$$\frac{x}{\begin{vmatrix} b_1 & c_1 & d_1 \\ b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \end{vmatrix}}=\frac{-y}{\begin{vmatrix} a_1 & c_1 & d_1 \\ a_2 & c_2 & d_2 \\ a_3 & c_3 & d_3 \end{vmatrix}}=\frac{z}{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}=\frac{-1}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}.$$

Proof. Let us use capital letters to denote the co-factors of the corresponding small letters in Δ. Any common solution of the given system of equations is also a solution of the equation

$$A_1(a_1x+b_1y+c_1z+d_1)+A_2(a_2x+b_2y+c_2z+d_2)$$
$$+A_3(a_3x+b_3y+c_3z+d_3)=0 \qquad ...(i)$$

The coefficient of x in the left hand side of (i) is

$$A_1a_1+A_2a_2+A_3a_3=\Delta,$$

the coefficient of y is

$$A_1b_1+A_2b_2+A_3b_3=0,$$

the coefficient of z is

$$A_1c_1+A_2c_2+A_3c_3=0,$$

and the constant term is

$$A_1d_1+A_2d_2+A_3d_3=\begin{vmatrix} b_1 & c_1 & d_1 \\ b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \end{vmatrix}.$$

Equation (i) may, therefore, be written as

$$\Delta x+\begin{vmatrix} b_1 & c_1 & d_1 \\ b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \end{vmatrix}=0. \qquad ...(ii)$$

Similarly we can show that a common solution of the given system is also a solution of the equations

$$\Delta y+\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}=0, \qquad ..(iii)$$

and

$$\Delta z+\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}=0. \qquad ...(iv)$$

Since $\Delta \neq 0$, it follows from (*ii*), (*iii*) and (*iv*) that the given system has a unique solution, namely

$$\frac{x}{\begin{vmatrix} b_1 & c_1 & d_1 \\ b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \end{vmatrix}}=\frac{-y}{\begin{vmatrix} a_1 & c_1 & d_1 \\ a_2 & c_2 & d_2 \\ a_3 & c_3 & d_3 \end{vmatrix}}=\frac{z}{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}=\frac{-1}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}$$

Example 16. *Solve the following system of linear equations by means of determinants :*

$$x+y+z=7,\ x+2y+3z=16,\ x+3y+4z=22.$$

Solution. Write the equations in the form

$$x+y+z-7=0,$$
$$x+2y+3z-16=0,$$
$$x+3y+4z-22=0.$$

Now $\Delta=\begin{vmatrix} 1 & 1 & 1 \\ 2 & 2 & 3 \\ 3 & 3 & 4 \end{vmatrix}=\begin{vmatrix} 1 & 0 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 3 \end{vmatrix}=\begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix}=-1.$

Since $\Delta \neq 0$, the system has a unique solution given by

$$\frac{x}{\begin{vmatrix} 1 & 1 & -7 \\ 1 & 3 & -16 \\ 1 & 4 & -22 \end{vmatrix}}=\frac{-y}{\begin{vmatrix} 1 & 1 & -7 \\ 1 & 3 & -16 \\ 1 & 4 & -22 \end{vmatrix}}=\frac{z}{\begin{vmatrix} 1 & 1 & -7 \\ 1 & 2 & -16 \\ 1 & 3 & -22 \end{vmatrix}}=\frac{-1}{\Delta}.$$

Now $\begin{vmatrix} 1 & 1 & -7 \\ 2 & 3 & -16 \\ 3 & 4 & -22 \end{vmatrix}=\begin{vmatrix} 1 & 0 & 0 \\ 2 & 1 & -2 \\ 3 & 1 & -1 \end{vmatrix}=\begin{vmatrix} 1 & -2 \\ 1 & -1 \end{vmatrix}=1,$

$$\begin{vmatrix} 1 & 1 & -7 \\ 1 & 3 & -16 \\ 1 & 4 & -22 \end{vmatrix}\quad\begin{vmatrix} 1 & 0 & 0 \\ 1 & 2 & -9 \\ 1 & 3 & -15 \end{vmatrix}=\begin{vmatrix} 2 & -9 \\ 3 & -15 \end{vmatrix}=-3,$$

$$\begin{vmatrix} 1 & 1 & -7 \\ 1 & 2 & -16 \\ 1 & 3 & -22 \end{vmatrix}=\begin{vmatrix} 1 & 0 & 2 \\ 1 & 1 & -9 \\ 1 & 2 & -15 \end{vmatrix}=\begin{vmatrix} 1 & -9 \\ 2 & -15 \end{vmatrix}=3.$$

Therefore

$$\frac{x}{1}=\frac{-y}{-3}=\frac{z}{3}=\frac{-1}{-1}.$$

Hence $x=1$, $y=3$, $z=3$.

EXERCISE 1 (*j*)

Solve the following systems of linear equations by means of determinants and verify your answer in each case :

1. $x+y-z=1,$
   $8x+3y-6z=1,$
   $-4x-y+3z=1,$

2. $4x-3y+2z=8,$
   $3x-4y+5z=6,$
   $-6x+5y+7z=-1.$

3. $3x+2y+5z=32,$
$2x+5y+3z=31,$
$5x+3y+2z=27.$

4. $3x-3y+5z=11,$
$5x+2y-7z=-12,$
$-4x+3y+z=5.$

5. $8x-7y-5z=1,$
$-7x+5y+6z=-1,$
$12x-8y-11z=2.$

6. $x+y+z+1=0,$
$x+2y+3z+4=0,$
$x+3y+4z+6=0.$

7. $x+2y+3z=6,$
$2x+4y+z=7,$
$3x+2y+9z=14.$

8. $x-y-2z=3,$
$2x+y+z=5,$
$4x-y-2z=11.$

9. $2x+3y+4z=16,$
$3x+2y-5z=8,$
$5x-6y+3z=6.$

10. $x+4y-4z=5,$
$3x-2y+2z=14,$
$-10x+8y+z=6.$

1·22. ADJOINT OF A MATRIX

In this and the next section we shall apply our knowledge of determinants to computation of inverses of 3×3 matrices. Later on, this technique can be extended to compute the inverses of $n\times n$ matrices.

Definition 1·10. *Let* $A=[a_{ij}]$ *be any* 3×3 *matrix and let* $B=[b_{ij}]$ *be a* 3×3 *matrix such that* $b_{ij}=A_{ji}$ *(where* A_{ji} *is the co-factor of* a_{ji} *in det* A*). Then* B *is called the adjoint of* A *and is written as adj* A.

Theorem 1·12. *If* A *be any 3-rowed square matrix, then*

$$(adj\ A)A=(det\ A)I=A(adj\ A),$$

where I *is the 3-rowed unit matrix.*

Proof. Let *adj* $A=[b_{ij}]$, where $b_{ij}=A_{ji}$.

Since A, *adj* A, are 3-rowed square matrices, therefore both A(*adj* A), (*Adj* A) A exist and are of type 3×3.

Also, (i, j)th element of A (*adj* A)

$$=\sum_k a_{ik}b_{kj},$$

$$=\sum_k a_{ik}A_{jk}$$

$$=\begin{cases}0, \text{ if } i\neq j\\ det\ A, \text{ if } i=j.\end{cases}$$

Therefore, A(*adj* A)=(*det* A)I.

Similarly we may show that

$$(adj\ A)A=(det\ A)I.$$

Corollary. If A be a 3-rowed square matrix such that *det* A$\neq$0, then A is invertible and

$$A^{-1}=\frac{1}{det\ A}\,adj\ A.$$

Proof. Since *det* A$\neq$0, therefore, the matrix $\frac{1}{det\ A}$ *adj* A exists. Call it B.

Now $$AB=A\left(\frac{1}{det\ A}\,adj\ A\right),$$
$$=\frac{1}{det\ A}\,(A\ adj\ A).$$
$$=\frac{1}{det\ A}\,(det\ A)I,$$
$$=I.$$

Similarly BA=I.

Since AB=BA=I, therefore, A is invertible and

$$B=\frac{1}{det\ A}(adj\ A)=A^{-1}.$$

Remark. The above corollary gives a rather powerful method of determining the inverse of a square matrix.

Example 17. *Find the adjoint of the matrix*

$$\begin{pmatrix} 1 & -2 & 3 \\ 0 & 2 & -1 \\ -4 & 5 & 2 \end{pmatrix}.$$

Solution. The co-factors of the elements of the first column are 9, 19, -4 respectively. Therefore, the first row of the adjoint is (9, 19, -4). The co-factors of the elements of the second column are 4, 14, 1 respectively. Therefore, the second row of the adjoint is (4, 14, 1). The co-factors of the elements of the third column are 8, 3, 2 respectively. Therefore the third row of the adjoint is (8, 3, 2).

Hence the desired adjoint matrix is

$$\begin{pmatrix} 9 & 19 & -4 \\ 4 & 14 & 1 \\ 8 & 3 & 2 \end{pmatrix}$$

Remark. It is always advisable to check the computations regarding the adjoint by using the identity

$$(adj\ A)A=A(adj\ A)=|\,A\,|\,I.$$

Example 18. *Find the inverse of the matrix*

$$\begin{pmatrix} 3 & -1 & 1 \\ -15 & 6 & -5 \\ 5 & -2 & 2 \end{pmatrix}$$

Solution. If the given matrix be denoted by A, then as in Example 17,

$$\text{adj } A=\begin{pmatrix} 2 & 0 & -1 \\ 5 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix}$$

Also, $\quad |A|=3.2+(-1)5+1.0=1.$

Therefore, $\quad A^{-1}=\dfrac{1}{|A|}\text{ adj } A,$

$$=\begin{pmatrix} 2 & 0 & -1 \\ 5 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix}.$$

Verification. Let us denote by B the matrix

$$\begin{pmatrix} 2 & 0 & -1 \\ 5 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix}.$$

Now $\quad AB=\begin{pmatrix} 3 & -1 & 1 \\ -15 & 6 & -5 \\ 5 & -2 & 2 \end{pmatrix}\begin{pmatrix} 2 & 0 & -1 \\ 5 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix}$

$$=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}=I_3.$$

Similarly $BA=I_3$.

Since $AB=BA=I_3$, therefore B is the inverse of A.

The following theorem is sometimes useful for computing inverses.

Theorem 1.13. *If* A *and* B *are n-rowed invertible matrices, then* A^t *and* AB *are both invertible, and*

(*a*) $\quad (A^t)^{-1}=(A^{-1})^t.$

(*b*) $\quad (AB)^{-1}=B^{-1}A^{-1}.$

Proof. (*a*) Since $AA^{-1}=A^{-1}A=I_n$,
therefore, by the reversal law for transposes, we have

$$(AA^{-1})^t=(A^{-1}A)^t=I_n^t,$$

i.e. $$(A^{-1})^tA^t=(A^t)(A^{-1})^t=I_n.$$

By the definition of the inverse of a matrix, it is immediate that A^t is invertible and $(A^{-1})^t$ is the inverse of A^t.

(*b*) Since A and B are n-rowed invertible matrices, therefore, A^{-1} and B^{-1} are both n-rowed square matrices, and consequently $B^{-1}A^{-1}$ is also an n-rowed square matrix. Let us denote $B^{-1}A^{-1}$ by C. To show that C is the inverse of the n-rowed square matrix AB, we have to verify that

$$C(AB)=(AB)C=I_n.$$

Now, $$C(AB)=(B^{-1}A^{-1})(AB),$$
$$=B^{-1}(A^{-1}A)B,$$
$$=B^{-1}I_nB=B^{-1}B=I_n,$$

and $$(AB)C=(AB)(B^{-1}A^{-1}),$$
$$=A(BB^{-1})A^{-1},$$
$$=AI_nA^{-1},$$
$$=AA^{-1}=I_n.$$

Thus we find that
$$C(AB)=(AB)C=I_n.$$

Hence it follows that AB is invertible and that C is the inverse of AB.

i.e., $$(AB)^{-1}=B^{-1}A^{-1}.$$

The result (*b*) of the above theorem is usually referred to as the reversal law for inverses.

As an application of the above theorem, consider the following :

Example 19.

If $$F(\alpha)=\begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$G(\beta)=\begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix},$$

show that the inverse of the matrix $F(\alpha)G(\beta)$ *is* $G(-\beta)F(-\alpha)$.

Solution. By the reversal law for inverses,
$$[F(\alpha)G(\beta)]^{-1}=[G(\beta)]^{-1}[F(\alpha)]^{-1} \qquad ...(i)$$

We have now to compute $[G(\beta)]^{-1}$ and $[F(\alpha)]^{-1}$, and show that they are equal to $G(-\beta)$ and $F(-\alpha)$ respectively.

Now, by actual multiplication we find that
$$F(\alpha)F(\gamma)=F(\alpha+\gamma)$$

so that $$F(\alpha)F(-\alpha)=F(0)=I_3,$$

and consequently
$$[F(\alpha)]^{-1}=F(-\alpha). \qquad ...(ii)$$

Similarly, $$G(\beta)G(\gamma)=G(\beta+\gamma),$$

so that $$G(\beta)G(-\beta)=G(0)=I_3,$$

where $$[G(\beta)]^{-1}=G(-\beta). \qquad ...(iii)$$

From (*i*), (*ii*) and (*iii*) we find that the inverse of $F(\alpha)G(\beta)$ is $G(-\beta)F(-\alpha)$.

EXERCISE 1(*k*)

1. If $A=\begin{pmatrix} 2 & 3 \\ 5 & -2 \end{pmatrix}$, show that $A^{-1}=\frac{1}{19}A$. (*A.I.S.S.C.E., 1986*)

2. Given $A=\begin{pmatrix} 2 & -3 \\ -4 & 7 \end{pmatrix}$, compute A^{-1} and show that $2A^{-1}=9I-A$. (*A.I.S.S C.E., 1987*)

Find the adjoint of each of the following matrices :

3. $\begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & -3 \\ 2 & -1 & 3 \end{pmatrix}$
4. $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 0 \\ 3 & -1 & 4 \end{pmatrix}$.
5. $\begin{bmatrix} 2 & 1 & 0 \\ 0 & -3 & 1 \\ -1 & -1 & 3 \end{bmatrix}$.
6. $\begin{pmatrix} -1 & -2 & 3 \\ -2 & 1 & 1 \\ 4 & -5 & 2 \end{pmatrix}$.

Find the inverse of each of the following matrices by first computing the adjoint :

7. $\begin{pmatrix} 1 & -2 & 3 \\ 0 & -1 & 4 \\ -2 & 2 & 1 \end{pmatrix}$.
8. $\begin{pmatrix} 2 & -1 & 4 \\ -3 & 0 & 1 \\ -1 & -1 & 2 \end{pmatrix}$.
9. $\begin{pmatrix} 2 & 1 & 2 \\ 2 & 2 & 1 \\ 1 & 2 & 2 \end{pmatrix}$.
10. $\begin{pmatrix} 4 & -5 & 6 \\ -1 & 2 & 3 \\ -2 & 4 & 7 \end{pmatrix}$.
11. $\begin{pmatrix} 2 & -2 & 4 \\ 2 & 3 & 2 \\ -1 & 1 & -1 \end{pmatrix}$
12. $\begin{pmatrix} 1 & -2 & 3 \\ 0 & 2 & -1 \\ -4 & 5 & 2 \end{pmatrix}$.
13. $\begin{pmatrix} 0 & 1 & -1 \\ 4 & -3 & 4 \\ 3 & -3 & 4 \end{pmatrix}$.
14. $\begin{pmatrix} 4 & 3 & 3 \\ -1 & 0 & -1 \\ -4 & -4 & -3 \end{pmatrix}$.
15. $\begin{pmatrix} 1 & 4 & 2 \\ 2 & 9 & 4 \\ 2 & 8 & 5 \end{pmatrix}$.
16. $\begin{pmatrix} 29 & 2 & 18 \\ 14 & 1 & 9 \\ 14 & 1 & 10 \end{pmatrix}$.

17. Find A^{-1}, if $A=\begin{pmatrix} 3 & -10 & -1 \\ -2 & 8 & 2 \\ 2 & -4 & -2 \end{pmatrix}$. (*A.I.S.S.C.E., 1988*)

18. Compute the inverse of the matrix

$$\begin{pmatrix} 1 & 2 & -2 \\ -1 & 3 & 0 \\ 0 & -2 & 1 \end{pmatrix}.$$ (*D.B.S.S.C.E., 1988*)

19. If $A=\begin{pmatrix} 1 & 2 & 5 \\ 2 & 3 & 1 \\ -1 & 1 & 1 \end{pmatrix}$, find A^{-1} and verify your result. (*A.I.S.S.C.E., 1985*)

20. Find the inverse of the matrix

$$\begin{pmatrix} 1 & 1 & 2 \\ 3 & 1 & -3 \\ 2 & 1 & -1 \end{pmatrix}.$$

1·23 APPLICATION OF MATRICES TO SOLUTION OF SYSTEMS OF LINEAR EQUATIONS

Consider the following systems of equations :

(α) $2y+3z=12,$
$4y+6z=25.$

(β) $2y+3z=12,$
$3y+2z=13.$

(γ) $2y+3z=12,$
$4y+6z=24.$

Let us first of all try to solve (α). Multiplying both sides of the first equation by -2 and adding it to the second equation, we get

$$0\,.\,y+0\,.\,z=1,$$

i.e.,
$$0=1,$$

which is clearly impossible. We conclude that *the given system* (α) *does not possess a solution, i.e.*, it is inconsistent.

Let us now try to solve the system (β). Multiplying both sides of the first equation by $-3/2$ and adding them to corresponding sides of the second equation, we have

$$\left\{\left(-\frac{3}{2}\right).\,2+3\right\}y+\left\{\left(-\frac{3}{2}\right).\,3+2\right\}z=\left(-\frac{3}{2}\right)12+13,$$

i.e.,
$$-\frac{5}{2}z=-5,$$

or
$$z=2.$$

Substituting $z=2$ in the first equation, we have $y=3$.

We find that if there is a solution, it is given by $y=3, z=2$. By actual substitution we find that it is indeed a solution of the given system. We conclude that *the system* (β) *possesses a unique solution.*

Let us now examine the system (γ). Multiplying the first equation throughout by -1 and adding to the second equation, we have

$$0\,.\,y+0\,.\,z=0,$$

which is always satisfied. The system (γ) is, therefore, equivalent to the single equation

$$2y+3z=12.$$

For each value of z, the above equation determines a value of y, which is given by

$$y=-\frac{3}{2}z+6.$$

Therefore *the system* (γ) *has infinitely many solutions*

$$y=-\frac{3}{2}k+6,$$

$$z=k,$$

k being any real number.

From the above discussion we find that given a system of *two* linear equations in *two* unknowns, there can be three possibilities :

(*i*) the system is inconsistent (no solution) ;

(*ii*) the system has a unique solution ;

(*iii*) the system has infinitely many solutions.

If instead of two linear equations in two variables we take m linear equations in n variables, even then these are the only three possibilities. We shall discuss these possibilities for the case of linear equations in 3 variables. The treatment is, however, the same in the general case as well.

In view of the above discussion we are faced with three questions :

Question 1. How to test whether a given system of linear equations is inconsistent?

Question 2. How to test whether a given system of linear equations has a unique solution, and in case the system has a unique solution, how to find it?

Question 3. How to test whether a given system of linear equations has infinitely many solutions, and in case it has infinitely many solutions, how to find them. All the above questions are related, and can be handled together.

We shall use matrices to discuss the questions raised above and therefore let us first of all see as to how we can express a given system of equations in matrix notation.

The system of equations (α) above can be written as

$$\begin{pmatrix} 2 & 3 \\ 4 & 6 \end{pmatrix}\begin{pmatrix} y \\ z \end{pmatrix}=\begin{pmatrix} 12 \\ 25 \end{pmatrix}.$$

Similarly, the system of equations (β) can be written as

$$\begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} y \\ z \end{pmatrix}=\begin{pmatrix} 12 \\ 13 \end{pmatrix},$$

and the system of equations (γ) can be written as

$$\begin{pmatrix} 2 & 3 \\ 4 & 6 \end{pmatrix}\begin{pmatrix} y \\ z \end{pmatrix}=\begin{pmatrix} 12 \\ 24 \end{pmatrix}.$$

More generally, the system of m equations

$$a_{11}x_1+a_{12}x_2+\dots+a_{1n}x_n=b_1$$
$$a_{21}x_1+a_{22}x_2+\dots+a_{2n}x_n=b_2$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$a_{m1}x_1+a_{2m}x_2+\dots+a_{mn}x_n=b_n$$

can be written as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \cdots & & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

or as AX=B,

where $A=[a_{ij}]$ is an $m\times n$ matrix,

$X=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ is an $n\times 1$ matrix, and

$B=\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$ is an $n\times 1$ matrix

Illustration. The system of equations

$$x-4y+7z=8,$$
$$3x+8y-2z=6,$$
$$7x-8y+26z=31,$$

can be written in matrix notation as

$$\begin{pmatrix} 1 & -4 & 7 \\ 3 & 8 & -2 \\ 7 & -8 & 26 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 8 \\ 6 \\ 31 \end{pmatrix},$$

or as AX=B,

where A is the 3×3 matrix $\begin{pmatrix} 1 & -4 & 7 \\ 3 & 8 & -2 \\ 7 & -8 & 26 \end{pmatrix}$,

X and B are the 3×1 matrices

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \text{ and } \begin{pmatrix} 8 \\ 6 \\ 31 \end{pmatrix} \text{ respectively.}$$

Let us now turn our attention to the three questions raised above. As we have already said, a unified approach to all the three questions is possible. However, if we happen to know that A is a non-singular matrix, then the following theorem tells us that the system of equations AX=B has a unique solution.

Theorem 1·14. *Let $AX=B$ be a system of n linear equations written in matrix notation. If A is non-singular, then the equations have a unique solution which is given by $X=A^{-1}B$.*

Proof. Since A is non-singular, it is invertible, and A^{-1} exists.

Premultiplying both sides of $AX=B$ by A^{-1} we find that if the system has a solution, it must be given by

$$A^{-1}(AX)=A^{-1}B, \text{ i.e. } (A^{-1}A)X=A^{-1}B,$$

or
$$IX=A^{-1}B \quad \text{or} \quad X=A^{-1}B,$$

which shows that if there exists a solution it must be given by $X=A^{-1}B$.

Also, $X=A^{-1}B \Rightarrow AX=A(A^{-1}B)=(AA^{-1})B=IB=B$
$$\Rightarrow AX=B,$$
so that $X=A^{-1}B$ is indeed a solution.

Thus the system has a unique solution $X=A^{-1}B$.

Example 20. *Use the matrix method to solve the system of equations*

$$3x-7y=-4$$
$$5x+2y=7 \qquad \textit{(D.B.S.S.C.E., 1988)}$$

Solution. The given system of equations can be written as

$$\begin{pmatrix} 3 & -7 \\ 5 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} -4 \\ 7 \end{pmatrix}.$$

Since $\begin{vmatrix} 3 & -7 \\ 5 & 2 \end{vmatrix}=3.2-(-7).5=41\neq 0,$

therefore $\begin{pmatrix} 3 & -7 \\ 5 & 2 \end{pmatrix}$ is non-singular, and therefore invertible.

The given system has the unique solution given by

$$\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} 3 & -7 \\ 5 & 2 \end{pmatrix}^{-1}\begin{pmatrix} -4 \\ 7 \end{pmatrix} \qquad ...(i)$$

Now $\begin{pmatrix} 3 & -7 \\ 5 & 2 \end{pmatrix}^{-1}=\frac{1}{41}\begin{pmatrix} 2 & 7 \\ -5 & 3 \end{pmatrix}$

Therefore from (*i*), we have

$$\begin{pmatrix} x \\ y \end{pmatrix}=\frac{1}{41}\begin{pmatrix} 2 & 7 \\ -5 & 3 \end{pmatrix}\begin{pmatrix} -4 \\ 7 \end{pmatrix}$$
$$=\frac{1}{41}\begin{pmatrix} 41 \\ 41 \end{pmatrix}=\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The solution of the given system of equations is given by $x=1$, $y=1$.

Verification. Substituting $x=1$, $y=1$ in the given equations, we have

$$3x-7y=3.1-7.1=-4$$
$$5x+2y=5.1+2.1=7.$$

Example 21. *Compute A^{-1} for the following matrix A*

$$A=\begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 3 & 1 & 1 \end{pmatrix}.$$

Hence solve the system of equations

$$y+2z+8=0$$
$$x+2y+3z+14=0$$
$$3x+y+z+8=0.$$

(*A.I.S S.C.E., 1986*)

Solution. The given system of equations can be written as :

$$\begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 3 & 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} -8 \\ -14 \\ -8 \end{pmatrix}$$

Now $\quad |A|=\begin{vmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 3 & 1 & 1 \end{vmatrix}$

$$=-(1.1-3.3)+2(1.1-2.3),$$
$$=-1(-8)+2(-5),$$
$$=-2\neq 0,$$

so that A is non-singular and hence invertible.

The co-factors of the elements of the first row are -1, 8 and -5 respectively. The co-factors of the elements of the second row are 1, -6 and 3 respectively. The co-factors of the elements of the third row are -1, 2, and -1 respectively.

$$\therefore \quad \text{adj } A=\begin{pmatrix} -1 & 1 & -1 \\ 8 & -6 & 2 \\ -5 & 3 & -1 \end{pmatrix}.$$

$$A^{-1}=\frac{1}{|A|}\text{adj } A=-\frac{1}{2}\begin{pmatrix} -1 & 1 & -1 \\ 8 & -6 & 2 \\ -5 & 3 & -1 \end{pmatrix}.$$

The solution of the given system is

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}=A^{-1}\begin{pmatrix} -8 \\ -14 \\ -8 \end{pmatrix},$$

$$=-\tfrac{1}{2}\begin{pmatrix} -1 & 1 & -1 \\ 8 & -6 & 2 \\ -5 & 3 & -1 \end{pmatrix}\begin{pmatrix} -8 \\ -14 \\ -8 \end{pmatrix},$$

$$=\begin{pmatrix} -1 \\ -2 \\ -3 \end{pmatrix}.$$

Thus the solution is $x=-1$, $y=-2$, $z=-3$.

Verification. Substituting $x=-1$, $y=-2$, and $z=-3$ in the given equations we find that

$$y+2z+8=(-2)+2(-3)+8=0,$$
$$x+2y+3z+14=-1+2(-2)+3(-3)+14=0,$$
$$3x+y+z+8=3(-1)+(-2)+(-3)+8=0.$$

1·23·1. General method for solving a system of linear equations by using matrices

Suppose we wish to solve a system of three linear equations in three variables, say, x, y and z.

The *first step* is to consider two equations out of the given equations, say the first and the second, and eliminate x from these equations by adding a suitable multiple of the first equation to the second equation. We thus get a linear equation in the two variables y and z.

Similarly, by considering the first and the third of the given equations and eliminating x from them we can obtain another equation in the two variables y and z.

The *second step* is to consider the two equations obtained on the completion of step 1 above. These equations will form a system of one of the types (α), (β), (γ) discussed at the beginning of this section, and consequently we shall find that exactly one of the following three possibilities holds :

(*i*) the system has no solution ;

(*ii*) the system has a unique solution ;

(*iii*) the system has infinitely many solutions.

By using matrices, the above computation can be arranged in a systematic manner. In fact the various operations involved can be performed on the rows of A and B, because

(*i*) Interchange of two equations is equivalent to an interchange of two rows of A as also of corresponding rows of B. We shall denote by $R_i \leftrightarrow R_j$ the operation of interchanging the ith and jth rows of a matrix.

(*ii*) Multiplication of an equation throughout by a scalar k is equivalent to multiplying the elements of a row of A and also the elements of the corresponding rows of B by k. We shall denote by $R_i \to k\, R_j$ the operation of multiplication of the ith row of a matrix by k.

(*iii*) Adding to the jth equation k times the ith equation is equivalent to adding k times the ith row of A (resp. B) to the jth row of A (resp. B). We shall denote this operation by $R_j \to R_j - kR_i$.

The method described above can be well understood by studying the following examples. It will be seen that the case of unique solution is also covered by the above method.

Example 22. *Solve the system of equations :*

$$x+4y+7z=8$$
$$3x+8y-2z=6$$
$$7x-8y+26z=31.$$

Solution.

Let $A=\begin{pmatrix} 1 & -4 & 7 \\ 3 & 8 & -2 \\ 7 & -8 & 26 \end{pmatrix}$, $B=\begin{pmatrix} 8 \\ 6 \\ 31 \end{pmatrix}$

and assume that there exists a matrix

$$X=\begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

such that AX=B.

Then

$$\begin{pmatrix} 1 & -4 & 7 \\ 3 & 8 & -2 \\ 7 & -8 & 26 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 8 \\ 6 \\ 31 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} 1 & -4 & 7 \\ 0 & 20 & -23 \\ 7 & -8 & 26 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 8 \\ -18 \\ 31 \end{pmatrix} \quad [R_2 \to R_2-3R_1]$$

$$\Rightarrow \begin{pmatrix} 1 & -4 & 7 \\ 0 & 20 & -23 \\ 0 & 20 & -23 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 8 \\ -18 \\ -25 \end{pmatrix} \quad [R_3 \to R_3-7R_1]$$

$$\Rightarrow \begin{pmatrix} 1 & -4 & 7 \\ 0 & 20 & -23 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 8 \\ -18 \\ -7 \end{pmatrix} \quad [R_3 \to R_3-R_2]$$

We have reduced the coefficient matrix to triangular form.

Now

$$\begin{pmatrix} 1 & -4 & 7 \\ 0 & 20 & -23 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 8 \\ -18 \\ -7 \end{pmatrix}$$

$$\Rightarrow \quad x-4y+7z=8,$$
$$20y-23z=-18$$

and

$$0=-7.$$

Since the conclusion $0=-7$ is false. therefore, our assumption that for some X, AX=B, is also false. Consequently there is no X for which AX=B, that is, the given system of equations has no solution, *i.e.*, it is inconsistent.

Example 23. *Solve the system of linear equations*

$$x-2y+3z=6$$
$$3x+y-4z=-7$$
$$5x-3y+2z=5.$$

Solution.

$$\text{Let} \quad A=\begin{pmatrix} 1 & -2 & 3 \\ 3 & 1 & -4 \\ 5 & -3 & 2 \end{pmatrix},\ B=\begin{pmatrix} 6 \\ -7 \\ 5 \end{pmatrix}$$

and assume that there exists a matrix

$$X=\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

such that $\qquad AX=B.$

$$\text{Then} \begin{pmatrix} 1 & -2 & 3 \\ 3 & 1 & -4 \\ 5 & -3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 \\ -7 \\ 5 \end{pmatrix}$$

$$\Rightarrow \quad \begin{pmatrix} 1 & -2 & 3 \\ 0 & 7 & -13 \\ 5 & -3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 \\ -25 \\ 5 \end{pmatrix} \ [R_2 \to R_2 - 3R_1],$$

$$\Rightarrow \quad \begin{pmatrix} 1 & -2 & 3 \\ 0 & 7 & -13 \\ 0 & 7 & -13 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 \\ -25 \\ -25 \end{pmatrix} \ [R_3 \to R_3 - 5R_1]$$

$$\Rightarrow \quad \begin{pmatrix} 1 & -2 & 3 \\ 0 & 7 & -13 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 \\ -25 \\ 0 \end{pmatrix} \ [R_3 \to R_3 - R_2]$$

We have reduced the coefficient matrix A to triangular form.

$$\text{woN} \begin{pmatrix} 1 & -2 & 3 \\ 0 & 7 & -13 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 \\ -25 \\ 0 \end{pmatrix}$$

$$\Rightarrow \quad \begin{aligned} x-2y+3z &= 6, \\ 7y-13z &= -25, \\ 0 &= 0, \end{aligned}$$

$$\Rightarrow \quad x=\frac{13}{7}\ z=\frac{25}{7},$$

$$\begin{aligned} x &= 2y-3z+6, \\ &= 2\left(\frac{13}{7}z-\frac{25}{7}\right)-3z+6, \\ &= \frac{5}{7}z-\frac{8}{7}, \\ z &= z, \text{ for all } z. \end{aligned}$$

Thus we find that if (x, y, z) be a solution of the given system of equations, then it must be given by

$$x=\frac{5}{7}z-\frac{8}{7},$$

$$y=\frac{13}{7}z-\frac{25}{7},$$
$$z=z, \text{ for all } z.$$

The next step is to show that every such set of values is in fact a solution of the given system of equations.

By multiplication, we can easily check that

$$X=\begin{pmatrix} \frac{5}{7}z-\frac{8}{7} \\ \frac{13}{7}z-\frac{25}{7} \\ z \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & -2 & 3 \\ 3 & 1 & -4 \\ 5 & -3 & 2 \end{pmatrix} X=\begin{pmatrix} 6 \\ -7 \\ 5 \end{pmatrix}, \text{ for all } z.$$

Hence the complete solution of the given system of equations is

$$\left.\begin{aligned} x&=\frac{5}{7}z-\frac{8}{7}, \\ y&=\frac{13}{7}z-\frac{25}{7}, \\ z&=z. \end{aligned}\right\} \text{ for all } z.$$

i.e.
$$\left.\begin{aligned} x&=\frac{5}{7}k-\frac{8}{7}, \\ y&=\frac{13}{7}k-\frac{25}{7}, \\ z&=k. \end{aligned}\right\} \text{ for all } k.$$

Example 24. *Solve the system of linear equations*

$$2x-3y+4z=3,$$
$$x-3z=-2.$$

Solution : Let $A=\begin{pmatrix} 1 & -3 & 4 \\ 1 & 0 & -3 \end{pmatrix}$, $B=\begin{pmatrix} 3 \\ -2 \end{pmatrix}$

and assume that there exists a matrix

$$X=\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

such that $AX=B$.

Then $\begin{pmatrix} 2 & -3 & 4 \\ 1 & 0 & -3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 3 \\ -2 \end{pmatrix}$

$$\Rightarrow \begin{pmatrix} 1 & 0 & -3 \\ 2 & -3 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} -2 \\ 3 \end{pmatrix}, \; [R_1 \leftrightarrow R_2]$$

$$\Rightarrow \begin{pmatrix} 1 & 0 & -3 \\ 0 & -3 & 10 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} -2 \\ 7 \end{pmatrix} \quad [R_2 \to R_2-2R_1]$$

We have reduced the coefficient matrix A to triangular form.

$$\text{Now } \begin{pmatrix} 1 & 0 & -3 \\ 0 & -3 & 10 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} -2 \\ 7 \end{pmatrix}$$

$\Rightarrow$ $\qquad x-3z=-2,$

and $\qquad 3y+10z=7,$

$\Rightarrow$ $\qquad x=3z-2,\ \ y=\frac{10}{3}z-\frac{7}{3},\ z=z,$ for all z, that is,

$$x=3k-2,\quad y=\frac{10}{3}k-\frac{7}{3},\ z=k, \text{ for all } k.$$

It is easy to check that

$$\begin{pmatrix} 2 & -3 & 4 \\ 1 & 0 & -3 \end{pmatrix}\begin{pmatrix} 3k-2 \\ \frac{10}{3}k-\frac{7}{3} \\ k \end{pmatrix}=\begin{pmatrix} 3 \\ -2 \end{pmatrix}, \text{ for all } k.$$

Hence the complete solution of the given system of equations is

$$x=3k-2,\ y=\frac{10}{3}k-\frac{7}{3},\ z=k, \text{ for all } k.$$

Example 25. *Solve the system of equations*

$$3z-4y=2,$$
$$5x+2y=12,$$
$$-x+3y=1.$$

Solution. Let $A=\begin{pmatrix} 3 & -4 \\ 5 & 2 \\ -1 & 3 \end{pmatrix},\ B=\begin{pmatrix} 2 \\ 12 \\ 1 \end{pmatrix}$

and assume that there exists a matrix

$$X=\begin{pmatrix} x \\ y \end{pmatrix},$$

such that $AX=B$.

$$\text{Then } \begin{pmatrix} 3 & -4 \\ 5 & 2 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} 2 \\ 12 \\ 1 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} -1 & 3 \\ 5 & 2 \\ 3 & -4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} 1 \\ 12 \\ 2 \end{pmatrix}, \quad [R_1 \leftrightarrow R_3]$$

$$\Rightarrow \begin{pmatrix} -1 & 3 \\ 0 & 17 \\ 3 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 17 \\ 2 \end{pmatrix}, \ [R_2 \to R_2 + 5R_1]$$

$$\Rightarrow \begin{pmatrix} -1 & 3 \\ 0 & 17 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 17 \\ 5 \end{pmatrix}, \ [R_3 \to R_3 + 3R_1],$$

$$\Rightarrow \begin{pmatrix} -1 & 3 \\ 0 & 17 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 17 \\ 0 \end{pmatrix}, \ [R_3 \to R_3 - \tfrac{5}{17} R_2]$$

$$\Rightarrow \qquad x+3y=1, \qquad 17y=17$$

$$\Rightarrow \qquad x=2, \qquad y=1.$$

Thus we find that if the given system has a solution, then it must be given by $x=2$, $y=1$.

By actual multiplification we find that

$$\begin{pmatrix} 3 & -4 \\ 5 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 12 \\ 1 \end{pmatrix}$$

so that $x=2, y=1$ is in fact a solution.

Hence the complete solution of the given system of equations

$$x=2, \quad y=1.$$

Example 26. *Solve the system of equations*

$$x+y+z=17,$$
$$x+2y+4z=16,$$
$$x+3y+4z=22.$$

Solution. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 4 \end{pmatrix}, \ B = \begin{pmatrix} 7 \\ 16 \\ 22 \end{pmatrix}$

and assume that there exists a matrix

$$X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

such that $\qquad AX=B.$

Then $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ 16 \\ 22 \end{pmatrix}$

$$\Rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 3 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ 9 \\ 22 \end{pmatrix}, \ [R_2 \to R_2 - R_1],$$

$$\Rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ 9 \\ 22 \end{pmatrix}, [R_2 \to R_3 - R_1],$$

$$\Rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ 9 \\ -3 \end{pmatrix}, [R_3 \to R_3 - 2R_2].$$

we have reduced the co-efficient matrix A to triangular form.

$$\text{Now} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ 9 \\ -3 \end{pmatrix}$$

$$\Rightarrow \quad x+y+z=7, \qquad x+2z=9, \qquad -z=-3,$$

$$\Rightarrow \quad z=3, \qquad y=9-2z=3, \quad x=7-y-z=1.$$

Thus we find that if the given system possesses a solution then itmust be given by

$$x=1,\ y=3,\ z=3.$$

By multiplication we find that if

$$X = \begin{pmatrix} 1 \\ 3 \\ 3 \end{pmatrix}, \text{ then } \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 4 \end{pmatrix} X = \begin{pmatrix} 7 \\ 16 \\ 22 \end{pmatrix},$$

showing that X is in fact a solution of $AX=B$. Hence the given system of equations has a unique solution, namely

$$x=1,\ y=3,\ z=3.$$

Example 27. *Solve the systems of equations*

$$x+y-3z=0,$$
$$2x-y+2z=0,$$
$$3x-2y+z=0.$$

Solution : Let $A = \begin{pmatrix} 1 & 1 & -3 \\ 2 & -1 & 2 \\ 3 & -2 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

and assume that there exists a matrix

$$X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

such that $AX=B$.

$$\text{Then} \begin{pmatrix} 1 & 1 & -3 \\ 2 & -1 & 2 \\ 3 & -2 & 1 \end{pmatrix} X = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} 1 & 1 & -3 \\ 0 & -3 & 8 \\ 0 & -5 & 10 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

by the operations $R_2 \to R_2 - 2R_1$, $R_3 \to R_3 - 3R_1$.

$$\Rightarrow \begin{pmatrix} 1 & 1 & -3 \\ 0 & -5 & 10 \\ 0 & -3 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

by the operation $R_2 \leftrightarrow R_3$,

$$\Rightarrow \begin{pmatrix} 1. & 1 & -3 \\ 0 & 1 & -2 \\ 0 & -3 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

by the operation $R_2 \to -\frac{1}{5} R_2$.

$$\Rightarrow \begin{pmatrix} 1 & 1 & -3 \\ 0 & 1 & -2 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{cases} x+y-3z=0 \\ \quad y-2z=0, \\ \qquad 2z=0, \end{cases}$$

$$\Rightarrow \quad x=0,\ y=0,\ z=0,$$

Thus we find that if the given system possesses a solution, it must be given by

$$x=0,\ y=0,\ z=0,$$

Also, it is obvious that this is in fact a solution. Hence

$$x=y=z=0$$

is the only solution of the given system.

Remark. A linear equation of the form $ax+by+cz=0$ is called a homogeneous equation. If each equation of a given system of linear equations be homogeneous, then we say that the system of equations is a linear homogeneous system. The system of equations in the above example is thus a linear homogeneous system. It can be seen without any difficulty that every linear homogeneous system has a solution in which every variable takes the value zero. This solution is usually called the *trivial solution* of the system. The trivial solution may sometimes be the only solution of a linear homogeneous system (as in Example 27). This, however, need not always be the case. For example, the linear homogeneous system in Example 28 possesses solutions other than the trivial solution.

A linear system is said to be non-homogeneous if it is not homogeneous. The linear systems in examples 22-26 are all non-homogeneous. As regards existence of solutions, there is one important difference between homogeneous and non-homogeneous

systems. While a non-homogeneous system may not possess any solution (for example, the system in Example 22), every homogeneous system possesses at least one solution, namely, the trivial solution.

Example 28. *Solve the system of equations*

$$-2x+y-z=0,$$
$$x-2y+3z=0,$$
$$-7x+2y-z=0.$$

Solution. Let $A=\begin{pmatrix}-2 & 1 & -1\\ 1 & -2 & 3\\ -7 & 2 & -1\end{pmatrix}$,

and assume that there exists a matrix

$$X=\begin{pmatrix}x\\ y\\ z\end{pmatrix}$$

such that $AX=0$, where 0 stands for the null matrix of type 3×1.

Then $$\begin{pmatrix}-2 & 1 & -1\\ 1 & -2 & 3\\ -7 & 2 & -1\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=0,$$

$$\Rightarrow \begin{pmatrix}1 & -2 & 3\\ -2 & 1 & -1\\ -7 & 2 & -1\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=0 \qquad [R_2\leftrightarrow R_1]$$

$$\Rightarrow \begin{pmatrix}1 & -2 & 3\\ 0 & -3 & 5\\ 0 & -12 & 20\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=0$$

$$[R_2\to R_2-2R_1,\ R_3\to R_3\to R_1]$$

$$\Rightarrow \begin{pmatrix}1 & -2 & 3\\ 0 & -3 & 5\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=0 \qquad [R_3\to R_3-4R_2]$$

$$\Rightarrow \begin{cases}x-2y+3z=0,\\ -3y+5z=0.\end{cases}$$

$$\Rightarrow \begin{cases}x=\frac{1}{3}z,\\ y=\frac{5}{3}z.\end{cases}$$

By multiplication we find that if

$$x=\begin{pmatrix} \frac{1}{3}z \\ \frac{5}{3}z \\ z \end{pmatrix},$$

then AX=0, whatever value z may have.

Thus $x=\frac{1}{3}k$, $y=\frac{5}{3}k$, $z=k$, where k takes any value whatever, is the complete solution of the given system.

EXERCISE 1 (*l*)

Use matrix method to solve the following equations :

1. $3x+2y=0$, $x+y=1$. (*A.I.S.S.C.E.*, *1984*)
2. $5x+2y=4$, $7x+3y=5$. (*A.I.S.S.C.E.*, *1984*)
3. $x+y+z=6$, $x-y+z=2$, $2x+y-z=1$. (*A.I.S.S.C.E.*, *1987*)
4. $2x-y+4z=1$, $3x-z=2$, $x-y-2z=3$. (*A.I.S.S.C.E.*, *1987*)
5. $x+y+z=3$, $2x-y+z=2$, $x-2y+3z=2$. (*A.I.S.S.C.E.*, *1989*)
6. $8x+4y+3z=18$, $2x+y+z=5$, $x+2y+z=5$. (*A.I.S.S.C.E.*, *1988*)
7. $x-y+z=4$, $2x+y-3z=0$, $x+y+z=2$. (*D.B.S.S.C.E.*, *1989*)
8. $2x+8y+5z=5$, $x+y+z=-2$, $x+2y-z=2$. (*D.B.S.S.C.E.*, *1985*)

Which of the following systems of equations are consistent :

9. $x-3y+z=-1$,
$2x+y-4z=-1$,
$6x-7y+8z=-7$.

10. $2x-5y+7z=6$,
$x-3y+4z=3$,
$3x-8y+11z=11$.

11. $x+y+z=7$,
$x+2y+3z=16$,
$x+3y+4z=22$.

12. $x+y+z=2$,
$x+2y+3z=5$,
$x+3y+6z=11$,
$x+4y+10z=21$.

Obtain complete solution for such of the following sets of equations as are consistent :

13. $x-3y-8z=-10$,
$3x+y-4z=0$,
$2x+5y-6z=13$.

14. $x+y+z=3$,
$3x-5y+2z=8$,
$5x-3y+4z=14$.

15. $x+y+z+w=4$,
$x+y+z-w=2$,
$x-y+z-w=0$.

16. $x-y+2z=4$,
$3x+y+4z=6$,
$x+y+z=1$.

Solve completely the system of equations :

17. $2x-3y+z=0$,
$x+2y-3z=0$,
$4x-y-2z=0$.

18. $x+2y+3z=0$,
$2x+3y+4z=0$,
$7x+13y+9z=0$.

19. $4x+5y+6z=0,$
$5x+6y+7z=0,$
$7x+8y+9z=0.$

20. $x-3y+2z=0,$
$7x-21y+14z=0,$
$-3x+9y-6z=0.$

TEST YOUR UNDERSTANDING I

In each of the following problems four alternatives are given out of which only one is correct. Put a tick mark (✓) against the correct alternative.

1. The matrix $\begin{pmatrix} 2 & 6 & 1 \\ 1 & 9 & 2 \end{pmatrix}$ is of type

   (*a*) 2 (*b*) 3 (*c*) 3×2 (*d*) 2×3

2. If $A=\begin{pmatrix} 1 & 2 \\ 5 & 6 \end{pmatrix}$, $B=\begin{pmatrix} 3 & 5 \\ x & -4 \end{pmatrix}$, and

   $A+B=\begin{pmatrix} 4 & 7 \\ 6 & 2 \end{pmatrix}$, then the value of x is

   (*a*) -1 (*b*) 1 (*c*) 0 (*d*) 4.

3. A is a 2×3 matrix and AB is a 2×5 matrix. B must be a

   (*a*) 5×3 matrix (*b*) 3×3 matrix
   (*c*) 3×5 matrix (*d*) 5×5 matrix.

4. If $AB=0$, then

   (*a*) $A=0$, (*b*) $B=0$, (*c*) $A=0$ and $B=0$
   (*d*) the matrices A and B may be both non-null.

5. If $\begin{pmatrix} x & 6 \\ 8 & 4 \end{pmatrix}=0$, then the value of x must be

   (*a*) 4 (*b*) 12 (*c*) 6 (*d*) 8.

6. The matrix $\begin{pmatrix} 18 & x \\ 6 & 8 \end{pmatrix}$ is non-invertible.

   The value of x is

   (*a*) 24 (*b*) 6 (*c*) 8 (*d*) 18.

7. The value of the determinant

$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{vmatrix}$$

   is

   (*a*) 6 (*b*) 0 (*c*) 3 (*d*) 2.

8. The system of equations $2x+3y=5$, $3x-2y=0$ possesses

   (*a*) a unique solution (*b*) no solution
   (*c*) infinitely many solutions (*d*) two solutions.

9. The system of equations $2x+3y=5$, $10x+15y=50$
   (*a*) is consistent (*b*) is inconsistent
   (*c*) has infinitely many solutions
   (*d*) has a unique solution.

10. The system of equations $x+y+z=0$, $2x+3y+4z=0$, $kx+y-z=0$ has a nonzero solution. The value of k is
   (*a*) 1 (*b*) -1 (*c*) 0 (*d*) 2.

REVIEW EXERCISE I

1. Find $8A-2B$ if

$$A=\begin{pmatrix} 0 & 2 & 3 \\ 2 & 1 & 4 \end{pmatrix} \text{ and } B=\begin{pmatrix} 7 & 6 & 3 \\ 1 & 4 & 5 \end{pmatrix}.$$

2. If $A=\begin{pmatrix} 3 & -5 \\ -4 & 2 \end{pmatrix}$, find $A^2-5A-14\,I$, where I is the unit matrix of order 2. (*D.B.S.S.C.E., 1989*)

3. If $A=\begin{pmatrix} -1 & 1 & -1 \\ 3 & -3 & 3 \\ 5 & -5 & 5 \end{pmatrix}$, $B=\begin{pmatrix} 0 & 4 & 3 \\ 1 & -3 & -3 \\ -1 & 4 & 4 \end{pmatrix}$,

compute A^2B^2. (*D.B.S.S.C.E., 1985*)

4. Let A be the matrix $\begin{pmatrix} 3 & 8 \\ 2 & 1 \end{pmatrix}$. Find A^{-1} and verify that $A^{-1}=\frac{1}{13}A-\frac{4}{13}I$, where I is the 2×2 unit matrix.

(*A.I.S.S.C.E., 1984*)

5. Obtain the inverse of $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$. When will this matrix not have the inverse ? (*A.I.S.S.C.E., 1986*)

6. Find A (*adj* A) for the matrix

$$A=\begin{pmatrix} 1 & -2 & 3 \\ 0 & 2 & -1 \\ -4 & 5 & 2 \end{pmatrix}$$ (*D.B.S.S.C.E., 1984*)

7. Compute the inverse of the matrix

$$A=\begin{pmatrix} 1 & 2 & 5 \\ 2 & 3 & 1 \\ -1 & 1 & 1 \end{pmatrix}$$

and verify that $A^{-1}A=I$. (*A.I.S.S.C.E., 1984*)

8. If $A=\begin{pmatrix} 2 & 0 & -1 \\ 5 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix}$

prove that

$A^{-1}=A^2-6A+11I.$ (*A.I.S.S.C.E., 1989*)

9. Solve the system of equations

$$x+2y=4,\ 2x+5y=9,$$

using matrix method. (*A.I.S.S.C.E., 1985*)

10. Solve the matrix equation

$$\begin{pmatrix} 5 & 4 \\ 1 & 4 \end{pmatrix} X=\begin{pmatrix} 1 & -2 \\ 1 & 3 \end{pmatrix}$$

where X is a 2×2 matrix. (*D.B.S.S.C.E., 1984*)

11. Solve the following linear equations using matrix method

$$x+y+z=9,\ 2x+5y+7z=52,\ 2x+y-z=0.$$

(*A.I.S.S.C.E., 1988*)

12. Find the inverse of the matrix

$$\begin{pmatrix} 2 & -3 & 2 \\ 2 & 2 & -3 \\ -3 & 3 & 2 \end{pmatrix}.$$

Hence solve the system of equations

$$2x-3y+2z=1$$
$$2x+2y-3z=1$$
$$-3x+3y+2z=2.$$

(*A.I.S.S.C.E., 1986*)

13. Obtain the inverse of the matrices

$$\begin{pmatrix} 1 & p & 0 \\ 0 & 1 & p \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ q & 1 & 0 \\ 0 & q & 1 \end{pmatrix}$$

and hence that of the matrix

$$\begin{pmatrix} 1+pq & p & 0 \\ q & 1+pq & p \\ 0 & q & 1 \end{pmatrix}$$

[**Hint :** Use theorem 1·13].

14. A matrix A satisfies the equation $A^2=A$. Show that if n is a positive or negative integer, then

$$(I+A)^n=I+(2^n-1)\,A.$$

15. A non-singular matrix A has the property $A^tA=AA^t$, where A^t is the transpose of A. Prove that $A^tA^{-1}=A^{-1}A^t$. Prove also that if $B=A^{-1}A^t$, then BB^t is the identity matrix.

Find A^t, AA^t, A^{-1} and B when

$$A=\begin{pmatrix} 2 & -2 & 1 \\ 2 & 1 & -2 \\ 1 & 2 & 2 \end{pmatrix}.$$

16. For each real number x such that $-1 < x < 1$, let A (x) be the matrix

$$(1-x^2)^{-1/2}\begin{pmatrix} 1 & -x \\ -x & 1 \end{pmatrix}.$$

Show that A (x) A (y)=A (z),

$$\text{where } z=\frac{x+y}{1+xy}$$

Deduce that $[\mathrm{A}\,(x)]^{-1}=\mathrm{A}\,(-x)$.

SUMMARY

1. Two matrices $\mathrm{A}=[a_{ij}]$ and $\mathrm{B}=[b_{ij}]$ are said to be equal if (*i*) they are comparable, (*ii*) $a_{ij}=b_{ij}$ for each pair of subscripts i and j.
2. If $\mathrm{A}=[a_{ij}]$, $\mathrm{B}=[b_{ij}]$ are two matrices of the same type, their sum $\mathrm{C}=[c_{ij}]$ is a matrix of the same type such that $c_{ij}=a_{ij}+b_{ij}$.
3. Addition of matrices is associative as well as commutative.
4. If $\mathrm{A}=[a_{ij}]$ be an $m\times n$ matrix and k be any complex number, then $k\mathrm{A}$ is the $m\times n$ matrix whose (i, j)th element is $k\,a_{ij}$. The matrix $k\mathrm{A}$ is called the scalar multiple of A by k.
5. If A and B be comparable matrices, and k, l be any complex numbers, then the matrix $k\mathrm{A}+l\mathrm{B}$ is said to be a linear combination of the matrices A and B.
6. Let $\mathrm{A}=[a_{ij}]$, $\mathrm{B}=[b_{ij}]$ be $m\times n$ and $n\times p$ matrices respectively. The $m\times p$ matrix $[c_{ij}]$, where
$$c_{ij}=a_{i1}b_{1j}+a_{i2}b_{2j}+\ldots\ldots+a_{in}b_{nj}$$
is called the product of the matrices A and B and is denoted by AB.
7. Multiplication of matrices is associative but is not commutative.
8. Multiplication of matrices distributes itself over addition.
9. If $\mathrm{A}=[a_{ij}]$ be an $m\times n$ matrix, then the $n\times m$ matrix $\mathrm{B}=[b_{ij}]$ such that $b_{ij}=a_{ji}$ is called the transpose of A and is denoted by A$'$.
10. A square matrix A is said to be non-singular if $|\mathrm{A}|\neq 0$.
11. An n-rowed square matrix A is said to be invertible if there exists an n-rowed square matrix B such that
$$\mathrm{AB}=\mathrm{BA}=\mathrm{I}.$$
12. A square matrix A is invertible iff it is non-singular.
13. Let $\mathrm{A}=[a_{ij}]$ be any 3×3 matrix and let $\mathrm{B}=[b_{ij}]$ be a 3×3 matrix such that $b_{ij}=\mathrm{A}_{ji}$, where A_{ji} is the cofactor of a_{ji} in det A. Then B is called the adjoint of A.
14. If A is a square matrix such that
$$|\mathrm{A}|\neq 0, \text{ then}$$
$$\mathrm{A}^{-1}=\frac{1}{|\mathrm{A}|}\text{ adj. A.}$$
15. If the vertices of a triangle ABC are (x_1, y_1), (x_2, y_2), (x_3, y_3), then its area is
$$\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}.$$

16. If $\Delta \equiv \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \neq 0,$

then the system of linear equations

$$a_1x+b_1y+c_1z+d_1=0,$$
$$a_2x+b_2y+c_2z+d_2=0,$$
$$a_3x+b_3y+c_3z+d_3=0,$$

possesses a unique solution, which is given by

$$\frac{x}{\begin{vmatrix} b_1 & c_1 & d_1 \\ b_2 & c_2 & d_2 \\ b_3 & c_3 & d_3 \end{vmatrix}} = \frac{-y}{\begin{vmatrix} a_1 & c_1 & d_1 \\ a_2 & c_2 & d_2 \\ a_3 & c_3 & d_3 \end{vmatrix}} = \frac{z}{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}} = \frac{-1}{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}$$

17. If A is a square matrix which is non-singular, the matrix equation $AX=B$ has the unique solution $X=A^{-1}B$.

HISTORICAL NOTE

The word matrix was used for a rectangular array for the first time by *J J. Sylvester* (1814-1897) in 1850. Arthur Cayley (1821-1895) used matrices for the first time in 1855 to simplify notation in a paper on the study of invariants. Soon thereafter he wrote a paper entitled "A Memoir on the Theory of Matrices" in which he developed the basic concepts of matrices. He subsequently wrote several other papers on matrices. Cayley made such fundamental contributions to the theory of matrices that he is regarded as the creator of matrices. W.R. Hamilton made important contributions to matrices.

The theory of determinants was developed much before that of matrices. Some contributions to the theory of determinants were made by Leibnitz, Lagrange, Vandermonde, and Laplace. The first systematic account of the theory of determinants was, however, given by A.L. Cauchy (1789-1857) in 1812 in an eighty-four page long memoir. It was followed by several other papers by him. Jacobi also made contributions to the theory of determinants.

□ □

GUILLAUME FRANCOIS ANTOINE DE L'HOPITAL (1661-1704)

Guillaume Francois Antoine De L'Hopital was born in 1661. He was a French mathematician. Most of our undergraduate college calculus as discovered by Newton and Leibnitz had been consolidated by 1700. It was L 'Hopital who brought out the first printed textbook on the new calculus in 1696. The book was based on lectures of Johann Bernoulli who was a teacher of L'Hopital and who had written two small unpublished treatises on the differential and integral calculus during 1691-92. These were the results of Bernoulli which were compiled by L'Hopital in the form of his book entitled Analyse des infiniment petits pour l'intelligence des lignes courbes (Analysis of the infinitely small for the understanding of curves). The book contained basic formulae for differentials of algebraic functions with applications to problems involving tangents, maxima and minima, and curvature. However, this first calculus text of L'Hopital is now remembered mainly for its inclusion of a result concerning a familiar method of evaluating the indeterminate form 0/0. This result of Bernoulli became incorrectly known as L'Hopital's rule in later calculus texts and continues to be referred to as such even today.

CHAPTER 2

Real Functions, Limits and Continuity

2·1 REAL FUNCTIONS

A function f from a non-empty set S to a set T is a correspondence which assigns to each member x of S a member $f(x)$ of T. The set S is called the **domain** of f and the set $\{f(x) : x \in S\}$ is called the **range** of f. If S and T are subsets of **R**, the set of real numbers, then the function f is said to be a real function. Throughout the present book we shall deal with real functions only, unless stated otherwise.

In the present chapter we shall study some basic concepts such as the domain and range of a real function, graphs of real functions, limit of a real function, continuity of a real function etc.

The word 'function' will mean 'real function' throughout.

2·1·1. Domain and range of a real function

As you already know, the domain of a function is the set of values which the variable takes. If we are given only the functional relation, but the domain is not given explicitly, it is understood that the domain consists of all those real numbers which the variable can possibly take. The range is the set of all those values which the function takes as the variable takes all the values in the domain.

Illustrations

1. Let $f(x)=x^2$. Here x can take all real values, therefore the domain is **R**. As x varies over **R**, $f(x)$ takes all non-negative real values. Therefore the range is $R^+ \cup \{0\}$, the set of non-negative real numbers.

2. Let $f(x)=\sqrt{x}$. Here x can take all non-negative real values. Therefore domain of f is $R^+ \cup \{0\}$. As x varies over $R^+ \cup \{0\}$, $f(x)$ takes all non-negative real values. Therefore range of f is also $R^+ \cup \{0\}$.

Example 1. *Find the domain and range of each of the following functions :*

(*a*) $f(x)=\dfrac{1}{x^2-1}$ (*b*) $g(x)=\dfrac{x}{1+|x|}$

Solution.

(*a*) Since the denominator of $f(x)$ vanishes when $x=\pm 1$, therefore x cannot take the values -1 and 1. It can, of course, take every other value. Therefore D(f), the domain of $f=\mathbf{R}\sim\{-1, 1\}$.

Also, writing $y=\dfrac{1}{x^2-1}$, ...(1)

we have $$x=\pm\sqrt{\left(1+\frac{1}{y}\right)}.$$ ...(2)

For each of the two values of x given by (2), $f(x)=y$.

Therfore the range will consist of those real numbers y for which $\sqrt{\left(1+\dfrac{1}{y}\right)}$ is meaningful. Now $1+\dfrac{1}{y}=\dfrac{y+1}{y}=\dfrac{y(y+1)}{y^2}$, which is $\geqslant 0$ whenever $y>0$ and $y(y+1)\geqslant 0$, *i.e.* $y\geqslant -1$ or $y>0$. Therefore the range is the set $]-\infty, -1]\cup]0,\infty[$.

(*b*) The denominator of $g(x)$ does not vanish for any real value of x. Therefore x can take every value, and so D(g)=**R**.

Also, writing $$y=\frac{x}{1+|x|},$$ ...(1)

we find that $|y|<1$, *i.e. the range of g must be a subset of* $[-1, 1]$. ...(A)

From (1) we have

$$|y|=\frac{|x|}{1+|x|},$$

or $$1-|y|=\frac{1}{1+|x|}$$ ...(2)

From (1) and (2), we have

$$x=\frac{y}{1-|y|}$$ ...(3)

From (3), we find that if y be any real number such that $-1<y<1$, and x is given by (3), then $g(x)=y$. This shows that *every real number lying between* -1 *and* 1 *is in the range.* ...(B)

From (A) and (B) we find that

$$R(g)=]-1, 1[.$$

2·1·2. Sum, Difference, Product and Quotient of two real functions

It is possible to construct new functions by the four fundamental operations of addition, subtraction, multiplication and division on a given pair of real functions with a common domain.

Let f and g be two real functions with the same domain D.

(*a*) Define a function $s : D \to \mathbf{R}$ by setting

$$s(x) = f(x) + g(x), \forall x \in D.$$

The function s so defined is called *the sum of the functions* f and g, and is denoted by $f+g$.

Thus, $\forall x \in D$,

$$\boxed{(f+g)(x) = f(x) + g(x)}$$

(*b*) Define a function $d : D \to \mathbf{R}$ by setting

$$d(x) = f(x) - g(x), \ \forall x \in D.$$

The function d so defined is called *the function obtained by subtracting g from f*, and is denoted by $f-g$, Thus, $\forall x \in D$,

$$\boxed{(f-g)(x) = f(x) - g(x)}$$

(*e*) Define a function $p : D \to \mathbf{R}$ by setting

$$p(x) = f(x)\, g(x), \forall x \in D.$$

The function p is called the *product* of the functions f and g, and is denoted by fg. Thus, $\forall x \in D$,

$$\boxed{(fg)(x) = f(x) g(x)}$$

(*d*) Since division by zero is not permitted, therefore in order to be able to talk of the quotient of f by g we shall have to assume that $q(x) \neq 0$ for any $x \in D$. With this hypothesis, we define a function $q : D \to R$ by setting

$$q(x) = f(x)/g(x), \forall x \in D.$$

The function q is called the *quotient* of f by g, and is denoted by f/g. Thus, $\forall x \in D$,

$$(f/g)(x) = f(x)/g(x).$$

Remarks 1. In case $g(x) = 0$ for some $x \in D$, we can consider the set, say E, of all those $x \in D$ for which $g(x) \neq 0$, and define f/g on E by setting $(f/g)(x) = f(x)/g(x)$.

2. If in the above discussion the domains of f and g are not the same, then the functions $f+g, f-g, fg$ etc. can be defined on the intersection of the domains of f and g.

Illustration. Consider the functions

$$f : x \to x^4, \text{ and } g : x \to x^2, \forall x \in \mathbf{R}.$$

Then $f+g$, $f-g$ and fg are the functions

$$f+g : x \to x^4 + x^2, \forall x \in \mathbf{R}$$

$$f-g : x \to x^4 - x^2, \forall x \in \mathbf{R}$$
$$fg \quad : x \to x^6, \forall x \in \mathbf{R}$$

Since $g(x)=0 \Leftrightarrow x^2=0 \Leftrightarrow x=0$, therefore in order to define f/g we should consider only non-zero values of x.

If $x \neq 0$, then $(f/g)(x)=f(x)/g(x)=x^2$.

Thus f/g is the function

$$f/g : x \to x^2, \forall x \in \mathbf{R} \sim \{0\}.$$

Multiplication of a real function by a scalar

Let f be a real function with domain D, and let k be a fixed real number. It is sometimes useful to define a function kf : $\mathrm{D} \to \mathbf{R}$ by setting

$$(kf)(x)=k\,f(x), \ \forall x \in \mathrm{D}.$$

We sometimes say that kf is obtained from f by multiplying it with the scalar k.

Illustration. Let $f : x \to \sin x, \ \forall x \in \mathbf{R}$. Then $2f$ is the function $\mathbf{R} \to \mathbf{R}$ such that $(2f)(x)=2 \sin x, \ \forall x \in \mathbf{R}$.

Remark. Let g be the constant function

$$g(x)=k, \ \forall x \in \mathbf{R}.$$

Then kf is the same as the product of the functions g and f.

2·1·3. Composite of functions

We shall now describe a method of combining two functions which is different from the ones studied uptill now. In the preceding section we had considered functions with the same domain. Now we are going to consider a pair of functions such that the range of the first is contained in the domain of the second.

Let f and g be two functions with domain X and Y respectively, and let $f(\mathrm{X})$, $g(\mathrm{Y})$ be contained in Y and Z respectively. We define a function h on X by setting

$$h(x)=g(f(x)) \ \forall \ x \in \mathrm{X}.$$

To obtain $h(x)$ we first take the f-image of an element of X. Since $f(x) \in \mathrm{Y}$, therefore $f(x) \in \mathrm{Y}$. We now take the g-image of $f(x)$ so as to get $g(f(x))$ which is an element of $g(\mathrm{Y})$, the range of Y. This scheme has been shown in Fig. 2.1.

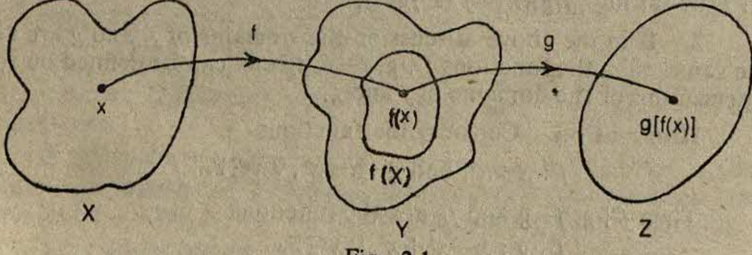


Fig. 2.1

The function h defined above is called the *composite* of f and g. We write it as $g \circ f$ (and read it as gohf). Note the order carefully—first f and then g to the left of it. We must distinguish it from $f \circ g$ which will be defined only if the range of g is contained in the domain of f.

Illustration.

Consider the functions

$$f : x \to x^3, \text{ for all } x \in \mathbf{R}$$
$$\text{and } g : x \to 3x+2, \text{ for all } x \in \mathbf{R}$$

$g \circ f$ is a function from $\mathbf{R}$ to itself defined by $(g \circ f)(x)$ $=g(f(x))=g(x^3)=3x^3+2$ for all $x \in$ R.

Also, $f \circ g$ is a function from $\mathbf{R}$ to itself defined by $(f \circ g)$ $(x)=f(g(x))=f(3x+2)=(3x+2)^3$ for all $x \in \mathbf{R}$.

Thus $g \circ f$ and $f \circ g$ are both defined but are different from each other.

The concept of composite of functions is used not only to combine functions but also to look upon a given function as made up of two or more simpler functions. For example, consider the function

$$h : x \to (5x-7)^4, \ \forall\ x \in \mathbf{R}.$$

We can think of h as made up of the functions

$$f : x \to 5x-7, \ \forall\ x \in \mathbf{R}$$
$$\text{and } g : u \to u^4, \ \forall\ u \in \mathbf{R}$$

If f and g be defined as above, then h is simply the function $g \circ f$.

One important fact about composite of functions is that two functions f: X→Y and g : Y→X are inverses of each other if and only if $g \circ f$ and $f \circ g$ are both identity functions, the first one with domain X and the second one with domain Y, *i.e.* $g \circ f=i_X$ and $f \circ g=i_Y$.

As an illustration, consider the functions

$$f : x \to 4x+5, \ \forall\ x \in \mathbf{R},$$
$$\text{and } g : x \to \frac{1}{4}x-\frac{5}{4}, \ \forall\ x \in \mathbf{R}.$$

It can be easily seen that $g \circ f$ and $f \circ g$ are both identity functions on $\mathbf{R}$, showing that f and g are a pair of inverse functions.

Example 2. *If $f(x)=(a-x^n)^{1/n}$, where $a>0$ and n is a positive integer, then show that $f[f(x)]=x$.* (*I.I.T. J.E.E. 1983*)

Solution. $f(f(x))=f[(a-x^n)^{1/n}]$

$$=\{a-[(a-x^n)^{1/n}]^n\}^{1/n}$$

$$=[a-(a-x^n)]^{1/n}$$
$$=x$$

Remark. $f \circ f$ is the identity function on **R**. Therefore

$$f=f^{-1}.$$

Example 3. *Let* $f(x)=\frac{x}{1+|x|}$ *for all* $x \in \mathbf{R}$, *and* $g(x)$ $=\frac{x}{1-|x|}$, *for all x such that* $-1<x<1$.

Find $g \circ f$ *and* $f \circ g$. *Also show that f is invertible, and find* f^{-1}.

Solution. *It can be easily seen* (*as in example 1, pape 7*) *that* $-1<f(x)<1$, *therefore* $f(x)$ *is in the domain of g. Therefore* $g(f(x))$: *has a meaning for all x in R.*

$$\text{Now } (g \circ f)(x)=g(f(x))$$
$$=g\left(\frac{x}{1+|x|}\right)$$
$$=\frac{\frac{x}{1+|x|}}{1-\left|\frac{x}{1+|x|}\right|}$$
$$=\frac{x}{1+|x|}(1+|x|)$$
$$=x, \text{ for all } x \in \mathbf{R}.$$

$\therefore$ $g \circ f$ is the identity function on **R**, the domain of f. ...(A)

Also, since the domain of f is **R**, therefore $f \circ g$ is defined for all x such that $-1<x<1$.

$$(f \circ g)(x)=f(g(x))$$
$$=f\left(\frac{x}{1-|x|}\right)$$
$$=\frac{\frac{x}{1-|x|}}{1+\left|\frac{x}{1-|x|}\right|}$$

Since we are considering only those x for which $|x|<1$, so that $1-|x|>0$, therefore

$$(f \circ g)(x)=\frac{x}{1-|x|}(1-|x|)=x.$$

Therefore $f \circ g$ *is the identity function on the domain of g.* ...(B)

From (A) and (B), we find that f is invertible and $f^{-1}=g$.

Remark. In the above example $g^{-1}=f$.

EXERCISE 2 (*a*)

Find the domain of each of the following functions :

1. (*a*) $f(x)=\frac{1}{x^2-2x}$ (*b*) $f(x)=\frac{1}{\sqrt{(x^2-3x)}}$
2. (*a*) $f(x)=\sqrt{|x|}$ (*b*) $f(x)=x^{1/3}$.
3. Find the range of each of the following functions :
   (*a*) $f(x)=4x+3$ (*b*) $f(x)=x^2+1$.
4. Find the domain and range of the function $f(x)=\frac{1}{5x+6}$.
5. Find the domain and range of the function $f(x)=\frac{2x+3}{3x+4}$.
   Is this function invertible ? If so, find its inverse.

2·2. GRAPHS OF REAL FUNCTIONS

We shall now give some examples of real functions and their graphs. These functions will be met with again and again, not only throughout this course, but also in your study of mathematics in the years to come.

(1) **The constant function.** Let f be the function defined on $\mathbf{R}$ by setting $f(x)=c$, for all $x \in \mathbf{R}$, where c is some fixed real number.

The graph of f is the straight line $y=c$. If $c>0$ it lies in the upper half of the xy-plane; if $c=0$, the graph of f is the x-axis; if $c<0$, the graph of f lies in the lower half of the xy-plane. In Fig. 2.2 we have shown the graphs of three constant functions.

(*a*) $f(x)=1 \ \forall x \in \mathbf{R}$ (Fig. 2·2 (*a*))

(*b*) $f(x)=0 \ \forall x \in \mathbf{R}$ (Fig. 2·2 (*b*))

(*c*) $f(x)=-1 \ \forall x \in \mathbf{R}$ (Fig. 2·2 (*c*))

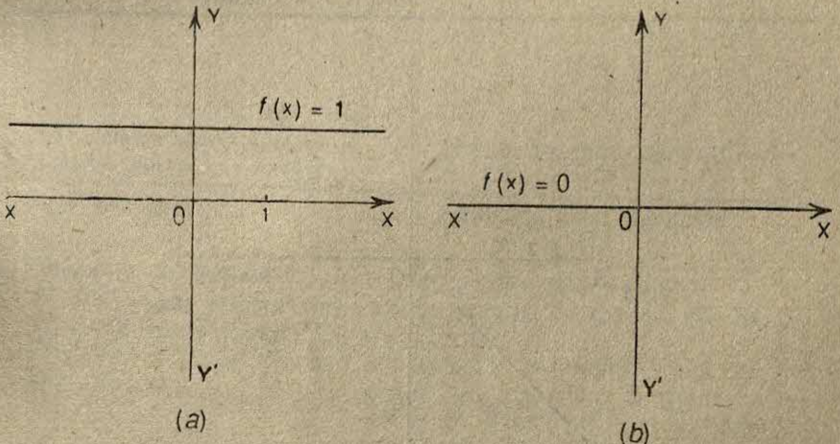
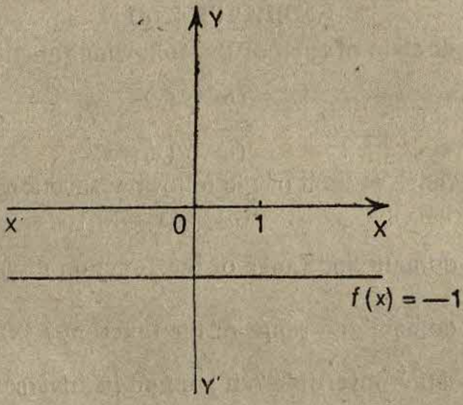


Fig. 2·2. Graphs of the functions defined by $f(x)=1$, $f(x)=0$ for all $x \in \mathbf{R}$.



(c)

Fig. 2·2. Graphs of the function defined by $f(x) = -1$, for all $x \in \mathbf{R}$.

By examining the nature of f we can obtain some important facts about the graph of f and *vice versa*. We tabulate below some properties of f and the corresponding facts about the graph of f.

Some facts about f	*Corresponding facts about the graph of f*
1. $f(x)$ is constant for all $x \in \mathbf{R}$	The graph of f is a straight line parallel to the x-axis
2. $f(x)=1$ for all $x \in \mathbf{R}$	The graph of f lies above the x-axis and is at a distance 1 unit from it.
3. $f(x)=0$ for all $x \in \mathbf{R}$	The graph of f is the x-axis.
4. $f(x)=-1$ for all $x \in \mathbf{R}$	The graph of f is below the x-axis and is at a distance -1.

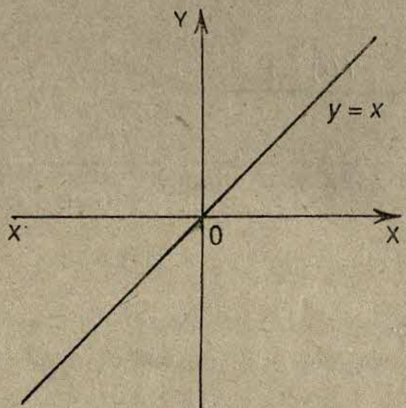


Fig. 2·3. Graph of the identity function

(2) **The identity function.** Let f be the function defined on **R** by setting $f(x)=x$, $\forall x \in \mathbf{R}$. The graph of f is the straight line $y=x$ (see Fig. 2·3).

(3) **The squaring function.** Let f be the function defined on **R** by setting $f(x)=x^2 \forall$, $x \in \mathbf{R}$.

The graph of f is the parabola $y=x^2$

	Some facts about f	*Corresponding facts about the graph of f*
1.	When $x=0$, $f(x)=0$	The graph passes through the origin.
2.	As x increases, $f(x)$ increases	The graph extends to infinity for $x \geqslant 0$.
3.	$f(-x)=f(x)$ for all $x \in \mathbf{R}$. *i.e.*, f is an even function of x	The graph is symmetric about the y-axis.

The graph is an shown in Fig. 2·4.

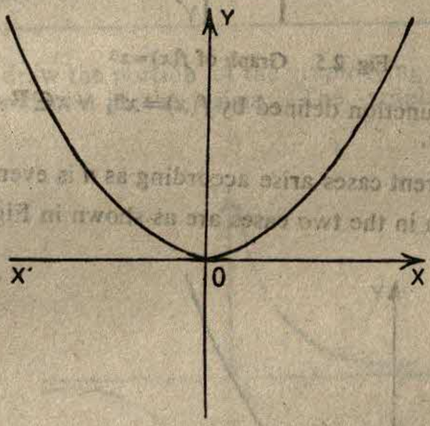


Fig. 2·4. Graph of $f(x)=x^2$

(4) **The cubing function.** Let f be the function defined on **R** by setting $f(x)=x^3$, $\forall x \in \mathbf{R}$.

	Some facts about f	*Corresponding facts about the graph of f*
1.	When $x=0$, $f(x)=0$	The origin lies on the graph.
2.	If $x>0$, then $f(x)>0$ and if $x<0$, then $f(x)<0$	The graph lies in the I and III quadrants.
3.	$f(-x)=-f(x)$, *i.e.*, f is an odd function	The graph is symmetric about the origin, *i.e.*, if (x, y) is a point on the graph, then $(-x, -y)$ is also a point on the graph.
4.	For $x>0$, if x increases then $f(x)$ increases	The graph extends to infinity in the first quadrant.

From the information gathered above we draw the portion of graph lying in the first quadrant and then reflect it in the origin to obtain the portion of the graph lying in the third quadrant. The graph is as shown in Fig. 2·5.

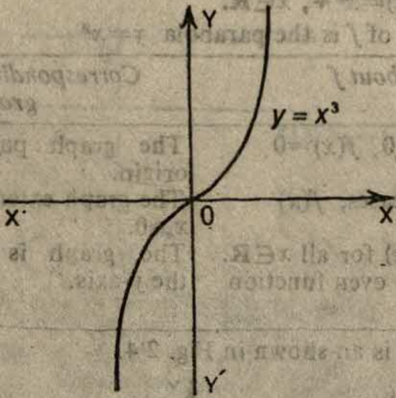


Fig. 2·5. Graph of $f(x)=x^3$

(5) The function defined by $f(x)=x^n$, $\forall x \in \mathbf{R}$ (n being a fixed positive integer).

Two different cases arise according as n is even or odd.

The graph in the two cases are as shown in Fig. 2·6.

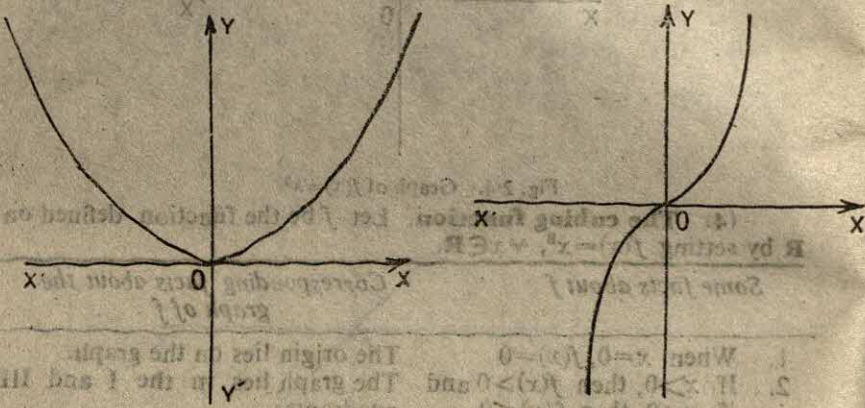


Fig. 2·6 (*a*) Graph of $f(x)=x^{2m}$ Fig. 2·6 (*b*) Graph of $f(x)=x^{2m+1}$

(6) **The reciprocal function.** Let f be defined by setting $f(x)=\dfrac{1}{x}$ whenever $x \neq 0$. The domain of f is $\mathbf{R} \sim \{0\}$.

Facts about f	*Corresponding facts about the graph of f*
1. f is not defined for $x=0$	The graph does not intersect the y-axis.
2. $f(x)\neq 0$ for any value of x	The graph does not intersect the x-axis.
3. f is an odd function	The graph is symmetric about the origin.
4. For $x>0$, as x takes values closer and closer to $0, f(x)$ becomes larger and larger	For small positive value of x, the points on the graph are close to the y-axis.
5. For $x>0$, as x becomes very large, $f(x)$ becomes very small.	For large positive values of x, the points on the graph are close to the x-axis.

We first draw the portion of the graph lying in the first quadrant and then reflect it in the origin. The graph is as shown in Fig. 2·7.

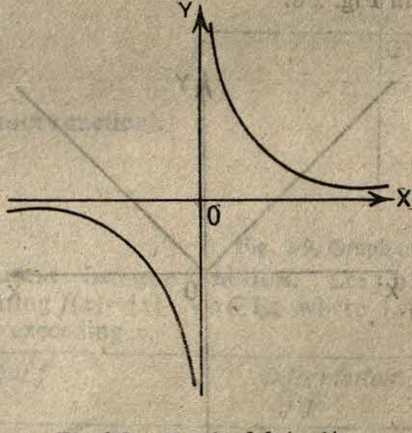


Fig. 2·7. Graph of $f(x)=1/x$.

(7) **The modulus function.** Let f be defined by setting $f(x)=|x|$ for all $x\in\mathbf{R}$. For this function, by definition we have

$$f(x)=\begin{cases} x, & \text{if } x\geqslant 0 \\ -x, & \text{if } x<0 \end{cases}$$

Facts about f	*Corresponding facts about the graph of f*
1. $f(x) \geqslant 0, \forall x \in \mathbf{R}$	The graph of f lies above the x-axis.
2. f is an even function of x	The graph is symmetric about the y-axis.
3. If $x \geqslant 0$, then $f(x) = x$.	The portion of the graph lying in the first quadrant consists of that part of the straight line $y = x$ which lies in the first quadrant.
4. If $x < 0$, then $f(x) = 0$.	The portion of graph lying in the second quadrant consists of that part of the straight line $y = -x$ which lies in the second quadrant.

Using the information obtained above we find that the graph of f is as shown in Fig. 2·8.

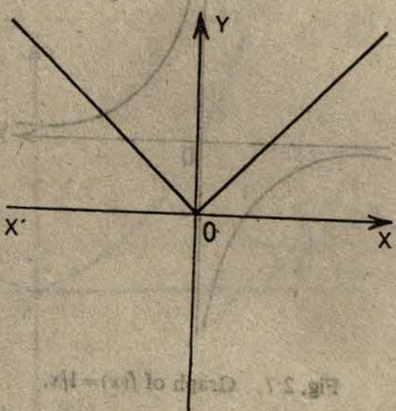


Fig. 2·8. Graph of $f(x) = |x|$

(8) Let f be the function defined for all $x \in \mathbf{R}$ by setting $f(x) = \sqrt{|x|}$.

Facts about f	Information about the graph of f
1. If $x=0$, then $f(x)=0$	The origin lies on the graph.
2. $f(x) \geqslant 0 \; \forall \; x \in \mathbf{R}$	The graph lies above the x-axis.
3. f is an even function of x.	The graph is symmetric about the y-axis. It is enough to draw the portion of the graph lying to the right of the y-axis. The graph can be completed by reflecting it in the y-axis.
4. If $x \geqslant 0$, then $f(x)=\sqrt{x}$	The portion of the graph lying in the first quadrant consists of the portion of the parabola $y^2=x$ lying in the first quadrant.

From the information gathered above, the graph of the function obtained as shown in Fig. 2 9.

Remark.

The function

$f(x)=\sqrt{|x|}$

for $x \geqslant 0$ is

called 'the square root function'.

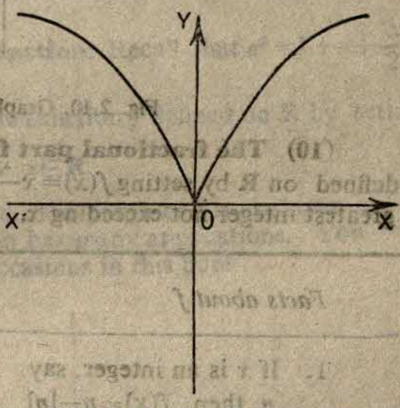


Fig. 2·9. Graph of $f(x)=\sqrt{|x|}$

(9) **The greatest integer function.** Let f be the function defined on **R** by setting $f(x)=[x] \; \forall x \in \mathbf{R}$, where $[x]$ denotes the greatest integer not exceeding x.

Facts about f	Information about the graph of f
1. If x is an integer, say $x=n$, then $f(x)=n$	The point (n, n) lies on the graph for all integers n.
2. If $n<x<n+1$, where n is an integer, then $f(x)=n$,	In the open interval $]n, n+1[$, the graph of f is the portion of the line $y=n$ lying between the ordinates $x=n$ and $x=n+1$.

From the above information we find that the graph consists of infinitely many steps of unit length rising by a unit height each time.

The graph is as shown in Fig. 2·10.

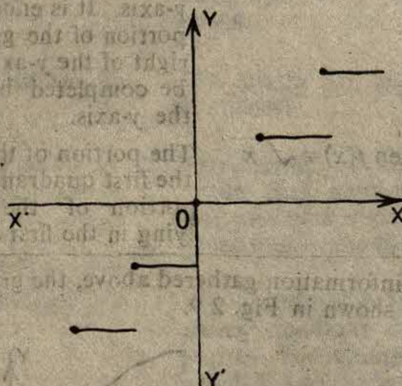


Fig. 2.10. Graph of $f(x)=[x]$.

(10) The fractional part function. Let f be the function defined on **R** by setting $f(x)=x-[x]$, $\forall x \in \mathbf{R}$, where $[x]$ denotes the greatest integer not exceeding x.

Facts about f	*Information about the graph of f*
1. If x is an integer, say n, then $f(x)=n-[n]$ $=n-n$ $=0.$	The point $(n, 0)$ lies on the graph for $n=0, \pm 1, \pm 2, \ldots\ldots$
2. If $n<x<n+1$, so that $x=n+q$, where $0<q<1$, then $f(x)=n+q-n=q$.	In the interval $]n, n+1[$, the graph is the portion of the straight line $y=x-n$ lying above the x-axis.
3. f is a periodic function with period 1.	The graph of f may be drawn when x lies in $[0, 1]$ and then repeated both ways by using periodicity.

The graph of f is as shown in Fig. 2·11.

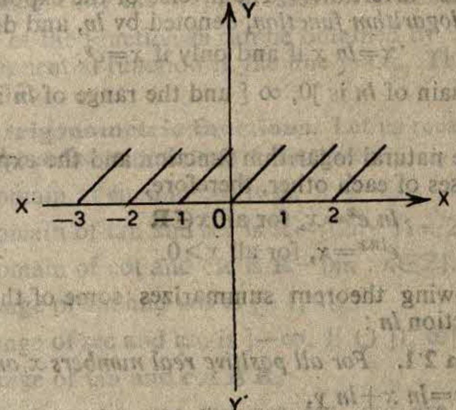


Fig. 2·11. Graph of $f(x)=x-[x]$

(11) The exponential function. Recall that $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{!3} + \text{.........}$, for all $x \in \mathbf{R}$. The function f defined on **R** by setting

$$f(x)=e^x, \ \forall \ x \in \mathbf{R}$$

is called the *exponential function.* The graph of f is as shown in Fig. 2.12. The exponential function has many applications. You will learn more about it on several occasions in this book.

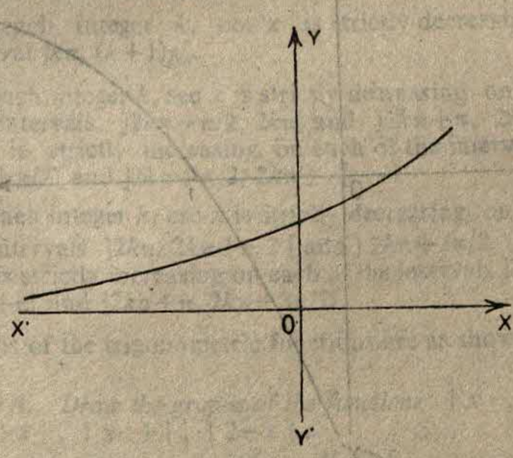


Fig. 2·12. Graph of $f(x)=e^x$

(12) **Logarithmic function**. Since the exponential function is a strictly increasing function with domain **R** and range]0, ∞], therefore it is invertible. The inverse of the exponential function is the natural *logarithm function*, denoted by *ln*, and defined by

$$x = \ln x \text{ if and only if } x = e^{y}.$$

The domain of *ln* is]0, ∞ [and the range of *ln* is the set R of real numbers.

Since the natural logarithm function and the exponential function are inverses of each other, therefore;

$$\ln e^{x} = x, \text{ for all } x \in \mathbf{R}$$
$$e^{\ln x} = x, \text{ for all } x > 0.$$

The following theorem summarizes some of the basic properties of the function *ln* :

Theorem 2·1. *For all positive real numbers x and y :*

(*i*) $\ln xy = \ln x + \ln y$,

(*ii*) $\ln \dfrac{x}{y} = \ln x - \ln y$,

(*iii*) $\ln x^{r} = r \ln x$, *for every* $r \in \mathbf{R}$,

(*iv*) $\ln x = \ln y$ if *and only if* $x = y$,

(*v*) $\ln x < \ln y$ if *and only if* $x < y$.

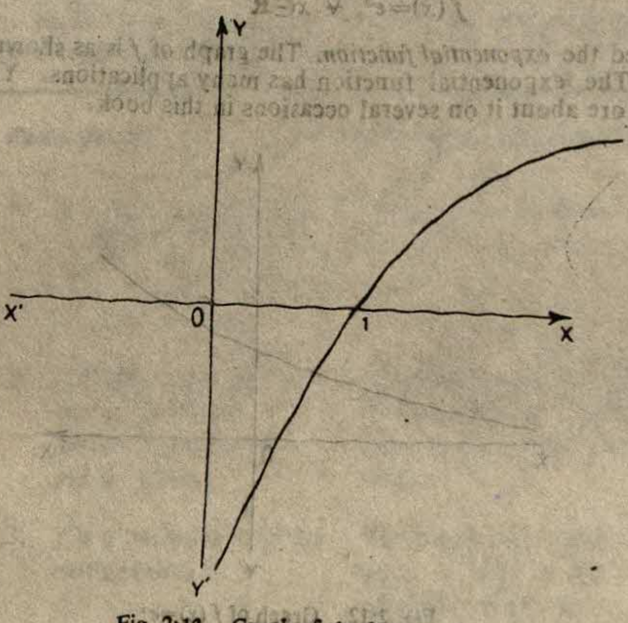


Fig. 2·13. Graph of the function *ln*

The graph of the function ln is wholly to the right of the y-axis since $ln\ x$ is defined only for $x>0$. Also, since $e^0=1$, therefore $ln\ 1=0$, and consequently the graph crosses the x-axis at (1, 0).

The graph of the function ln can be obtained by reflecting the graph of the exponential function in the line $y=x$. The graph is as shown in Fig. 2.13.

(13) **The trigonometric functions.** Let us recall the following basic facts about the trigonometric functions :

1. The domain of sin and cos is **R**.
2. The domain of tan and sec is $\mathbf{R}\sim\{(2n+1)\ \pi/2 : n\in\mathbf{Z}\}$.
3. The domain of cot and csc is $\mathbf{R}\sim\{n\pi : n\in Z\}$.
4. The range of sin and cos is $]-1, 1[$.
5. The range of sec and csc is $]-\infty, 1[\ \cup\ ]1, \infty[$.
6. The range of tan and cot is **R**.
7. The functions sin, cos, tan, cot, sec, csc are all periodic. The period of the functions sin, cos, sec and csc is 2π. The period of the functions tan and cot is π.
8. For each integer k, sin x strictly increases from -1 to $+1$ on the interval $[2k\pi-\pi/2, 2k\pi+\pi/2]$ and strictly decreases from $+1$ to -1 in the interval $[2k+\pi/2, 2k\pi+3\pi/2]$.
9. For each integer k, cos x strictly increases from -1 to $+1$ on the interval $[(2k-1)\pi, 2k\pi]$ and strictly decreases from $+1$ to -1 on the interval $[2k\pi, (2k+1)\pi]$.
10. For each integer k, tan x is strictly increasing on the interval $]k\pi-\pi/2, k\pi+\pi/2[$.
11. For each integer k, cot x is strictly decreasing on the interval $]k\pi, (k+1)\pi[$.
12. For each integer k, sec x is strictly decreasing on each of the intervals $]2k\pi-\pi/2, 2k\pi[$ and $]2k\pi+\pi, 2k\pi+3\pi/2[$ and is strictly increasing on each of the intervals $]\ 2k\pi, 2k\pi+\pi/2[$ and $]2k\pi+\pi/2, 2k\pi+\pi[$.
13. For each integer k, csc x is strictly decreasing on each of the intervals $]2k\pi, 2k\pi+\pi/2\ [$ and $]\ 2k\pi+3\pi/2, (2k+2)\pi[$, and is strictly increasing on each of the intervals $]2k\pi+\pi/2, 2k\pi+\pi[$ and $]2k\pi+\pi, 2k\pi+3\pi/2[$.

The graphs of the trigonometric functions are as shown in Fig. 2.14.

Example 4. *Draw the graphs of the functions* $|x|, -|x|, |x|+2, 1-|x|, |x+1|, |2-x|$.

Solution. (*a*) Since $|x|=\begin{cases} x, \text{ if } x\geqslant 0 \\ -x, \text{ if } x<0, \end{cases}$

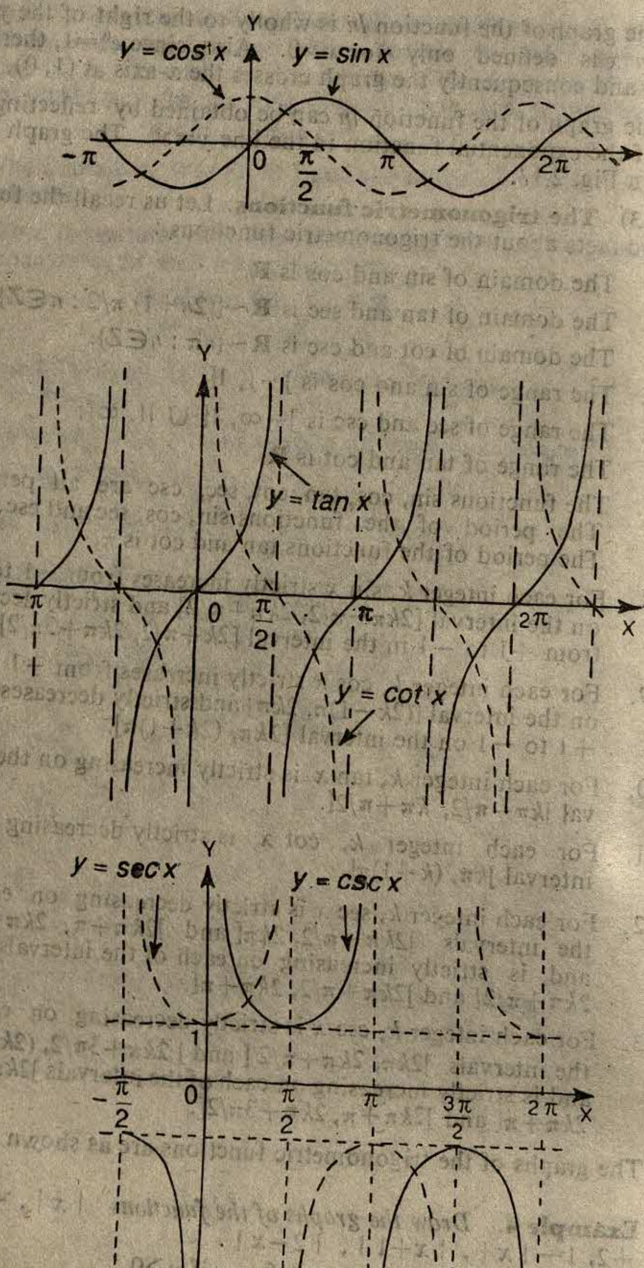
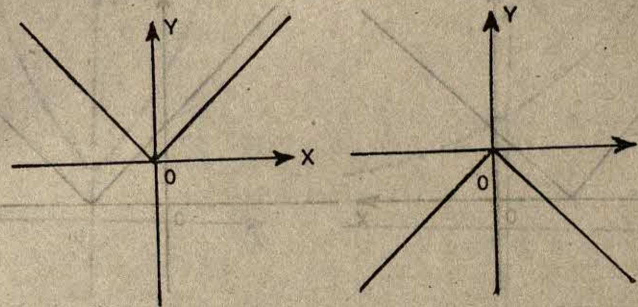


Fig. 2.14. Graphs of trigonometric functions.

therefore for $x \geqslant 0$ the graph consists of that portion of the straight line $y=x$ which lies in the right half-plane, and for $x<0$ the graph consists of that portion of $y=-x$ which lies in the left half-plane. The graph is as shown in Fig. 2.15 (*a*).

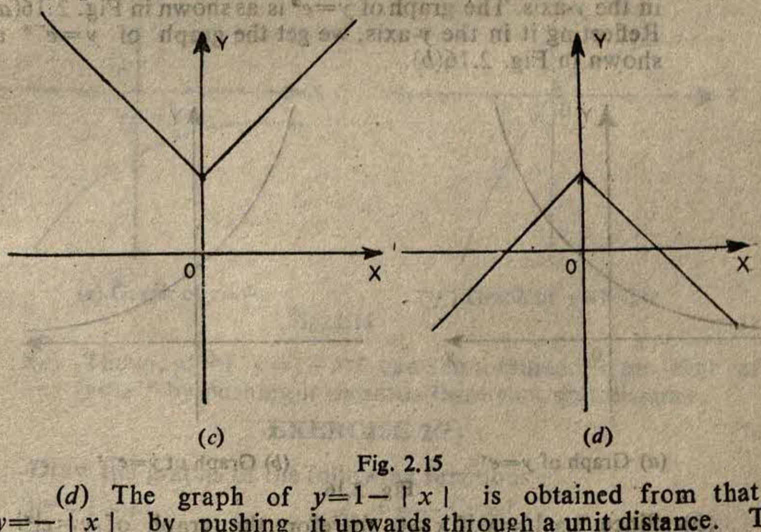


(*a*) Graph of $y=|x|$ (*b*) Graph of $y=-|x|$

Fig. 2.15

(*b*) Since (x, y) lies on $y=|x|$ iff $(x, -y)$ lies on $y=-|x|$, the graph of $y=-|x|$ is the reflection of the graph of $y=|x|$ in the *x*-axis (see Fig. 2.15 (*b*)).

(*c*) For each $x \in \mathbf{R}$, the ordinate of $y=|x|+2$ can be obtained by increasing the corresponding ordinate of $y=|x|$ by 2. Therefore the graph of $y=|x|+2$ can be obtained from that of $y=|x|$ by pushing it upward (*i.e.*, parallel to the *y*-axis by 2 units). The graph is as shown in Fig. 2.15 (*c*).

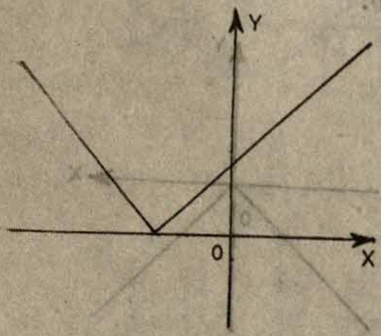


(*c*) (*d*)

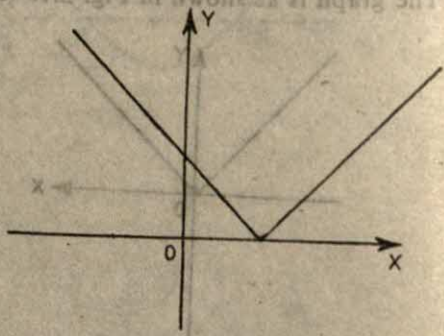
Fig. 2.15

(*d*) The graph of $y=1-|x|$ is obtained from that of $y=-|x|$ by pushing it upwards through a unit distance. The graph is as shown in fig. 2.15 (*d*).

(e) The graph of $y=|x+1|$ can be obtained from that of $y=|x|$ by pushing the graph to the left (*i.e.*, parallel to the x-axis) through a unit distance. The graph is as shown in fig. 2.15 (*e*).



(*e*) Graph of $y=|x|$



(*f*) Graph of $y=|x|$

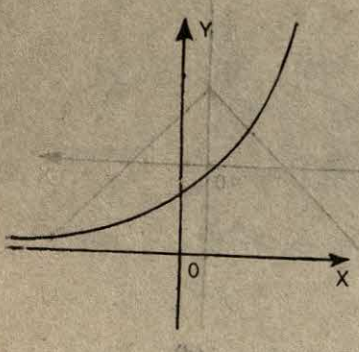
Fig. 2.15

(*f*) The graph of $y=|2-x|$ $(=|x-2|)$ can be obtained from that of $y=|x|$ by pushing the graph to the right (*i.e.*, parallel to the x-axis) through 2 units. The graph is as shown in fig. 2.15(*f*).

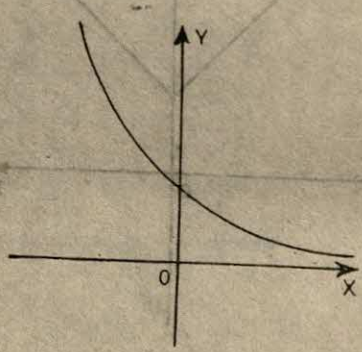
Example 5. *Draw the graphs of the functions* e^{-x}, $e^{|x|}$, $1+e^{x}$, $-e^{-x}$, $1-e^{-x}$.

Solution.

(*a*) The graph of $y=e^{-x}$ is the reflection of the graph of $y=e^{x}$ in the y-axis. The graph of $y=e^{x}$ is as shown in Fig. 2.16(*a*). Reflecting it in the y-axis, we get the graph of $y=e^{-x}$ as shown in Fig. 2.16(*b*).



(*a*) Graph of $y=e^{x}$

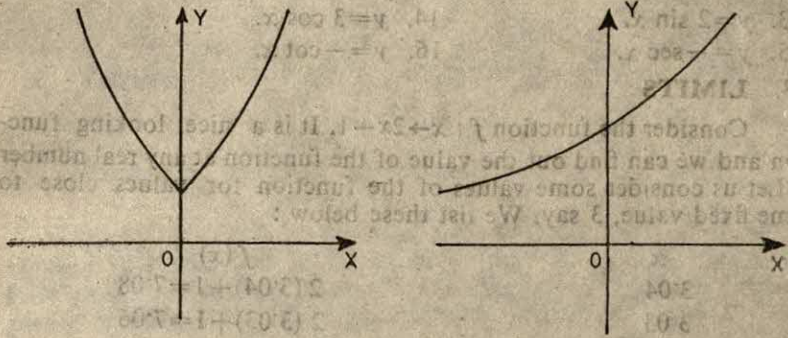


(*b*) Graph of $y=e^{-x}$

Fig. 2.16

(*b*) Since $|x|=x$ if $x\geqslant 0$, therefore the graph of $y=e^{|x|}$ is the same as that of $y=e^{x}$ if $x\geqslant 0$.

Since $|x|=-x$ if <0, therefore for $x<0$ the graph of $y=e^{|x|}$ is the same as that of $y=e^{-x}$. Using figures (*a*) and (*b*) above, the graph of $y=e^{|x|}$ can be drawn as shown in Fig. 2.16 (*c*).

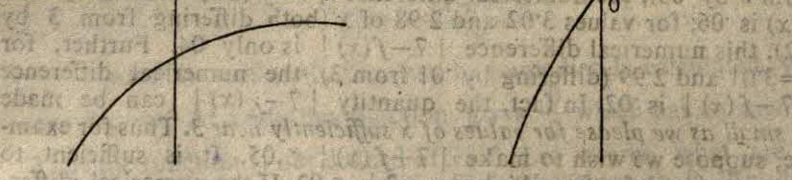


(*c*) Graph of $y=e^{|x|}$ (*d*) Graph of $y=1+e^{x}$

Fig. 2.16

Remark. We can also use the fact that the graph is symmetric about the y-axis.

(*d*) The graph is the reflection of that of $y=e^{-x}$ in the x-axis and is as shown in Fig. 2.16 (*e*).



(*e*) Graph of $y=-e^{-x}$ (*f*) Graph of $y=1-e^{-x}$.

Fig. 2.16

(*e*) The graph of $y=1-e^{-x}$ can be obtained from that of $y=e^{-x}$ by pushing it upwards through a unit distance.

EXERCISE 2(*b*)

Draw the graphs of the following functions :

1 $y=2|x|$. 2. $y=-3|x|$.

3. $y=2x^2$. 4. $y=1-x^2$.

5. $y=2\,|x-1|$. 6. $y=-3\,|x+2|$.
7. $y=|x-1|+|x+1|$. 8. $y=|x+2|-|x-2|$.
9. $y=2e^x$. 10. $y=-3e^{-x}$.
11. $y=-\ln x$. 12. $y=1+\ln x$.
13. $y=2\sin x$. 14. $y=3\cos x$.
15. $y=-\sec x$. 16. $y=-\cot x$.

2·3. LIMITS

Consider the function $f: x \rightarrow 2x+1$. It is a nice looking function and we can find out the value of the function at any real number x. Let us consider some values of the function for values close to some fixed value, 3 say. We list these below :

x	$f(x)$
3·04	$2(3·04)+1=7·08$
3·03	$2(3·03)+1=7·06$
3·02	$2(3·02)+1=7·04$
3·01	$2(3·01)+1=7·02$
2·99	$2(2·99)+1=6·98$
2·98	$2(2·98)+1=6·96$
2·97	$2(2·97)+1=6·94$
2·96	$2(2·96)+1=6·92$

If you look up the above table carefully, you will find that near $x=3$, values of $f(x)$ are near 7. For values of x very close to 3, values of $f(x)$ are very close to 7. As we take values of x (either greater or smaller than 3) which differ from 3 less and less, we find that the values of $f(x)$ differ less and less from 7.

Thus for example, when $x=3·03$ or $2·97$ (values of x differing from 3 by ·03), the numerical difference $|7-f(x)|$ between 7 and $f(x)$ is ·06; for values 3·02 and 2·98 of x (both differing from 3 by .02), this numerical difference $|7-f(x)|$ is only ·04. Further, for $x=3·01$ and 2·99 (differing by ·01 from 3) the numerical difference $|7-f(x)|$ is ·02. In fact, the quantity $|7-f(x)|$ can be made *as small as we please for values of x sufficiently near 3*. Thus for example, suppose we wish to make $|7-f(x)|<.05$. It is sufficient to choose values of x for which $|x-3|<.02$. If the numerical difference between x and 3 is less than .02 (such as for $x=3.001$, 3.002, 3·01, 2·999, 2·998, 2·99 etc.), then $|7-f(x)|$ would be less than ·05. This situation is described by saying that as x tends to 3, $f(x)$ tends to 7 or by saying that $f(x)$ *has limit* 7 *as* x *approaches* (*or tends to*) 3. We write it as

$$\lim_{x\rightarrow 3} f(x)=7.$$

The expression 'x tends to 3' is expressed symbolically as '$x\rightarrow 3$' and by this we mean that $|x-3|$ gets nearer and nearer zero without

actually being zero. We obtained the limiting value of $f(x)$ at $x=3$ and found it is 7. What is the actual value of $f(x)$ at $x=3$? By substitution, we have

$$f(3)=2\cdot3+1=7.$$

Thus the limiting value at $x=3$ is the same as the actual value. This need not be the case always, as is shown by the example below.

Example 6. *Consider the function* $f : \mathbf{R}\to\mathbf{R}$ *defined by*

$$f(x)=1, \text{ if } x\neq 0.$$
$$f(0)=0.$$

Evaluate $\lim_{x\to 0} f(x)$.

Solution. Since $f(x)=1$ for all non-zero x, therefore, for all values h of x sufficiently close to zero (but not actually zero), we have $f(h)=1$. Thus $|f(h)-1|=0$ (smaller than 'as small as we please' !) when h is sufficiently close to 0. Hence $\lim_{x\to 0} f(x)=1$.

But $f(0)=0$. Thus $\lim_{x\to 0} f(x)\neq f(0)$.

Remark. In the two examples that we discussed above, we calculated limits at points where the function had an actual value also. That, however, is not necessary in order to find out the limit of the function at a point. All that we must have is that the values of the function be known at all points sufficiently close to the point at which we wish to find the limit. We only need that function should be defined in some neighbourhood of the point except possibly at the point. (Whether the limit exists is a question to be discussed later !). Consider the following example as a clarification.

Example 7. *Find* $\lim_{x\to 0} f(x)$ *where*

$$f(x)=\frac{(x+1)^2-1}{x}.$$

Solution. If $x\neq 0$, $f(x)$ can be immediately calculated. What about $x=0$? At $x=0$, this function gives $\frac{0}{0}$ which has no meaning. Thus the function is not defined at $x=0$. Let us see what are the values of this function at points sufficiently close to 0. By direct calculations, we compute the following table :

x	$f(x)$
·05	2·05
·04	2·04
·03	2·03
·02	2.02
·01	2·01
−·01	1·99

$-\cdot 02$	$1\cdot 98$
$-\cdot 03$	$1\cdot 97$
$-\cdot 04$	$1\cdot 96$
$-\cdot 05$	$1\cdot 95$

A glance at the above table will show that as x takes values nearer and nearer 0, the values taken by $f(x)$ become closer and closer 2. In fact, by taking x sufficiently near x, the numerical difference between the value of $f(x)$ and 2 can made as small as we please. Thus for example, if we want that $|f(x)-2|$ should be smaller than 10^{-5}, we may take x so that $x \neq 0$ but $|x| < 10^{-5}$. Similarly for any other given number. Hence $\lim_{x \to 0} f(x)=2$. Thus we find that the function has a limit at $x=0$, even though the function is not defined for $x=0$.

In the examples that we have considered so far, the limit involved could always be found, whether or not the function was defined at that point. Only it was necessary that the function be defined in some neighbourhood of the point. This condition is, however, only necessary; it is not sufficient, meaning thereby that the function may be defined in some neighbourhood of a point and yet it may not have a limit at that point. The proof of this assertion is provided by the example that follows :

Example 8. Show that $\lim_{x \to 0} \left(\frac{1}{x}\right)$ does not exist.

Solution. For values of x other than zero, the function is well defined. Since we cannot divide by zero, so the function is not defined at $x=0$. As x takes smaller and smaller positive values, $1/x$ assumes greater and greater values. We cannot find any number l, such that $\left|\frac{1}{x}-l\right|$ will be as small as we please for values of suitably near 0. For suppose it was possible to find such a number l. Suppose we wish to make $\left|\frac{1}{x}-l\right| < \frac{1}{2}$. To be definite suppose $l > 0$. Since the numerical difference between l and $1/x$ is less than $\frac{1}{2}$, $1/x$ should be somewhere in between the points A and B, shown in the diagram below.

Fig. 2 17.

Thus we must have $\frac{1}{x} < l + \frac{1}{2} = \frac{(2l+1)}{2}$

$$\Rightarrow x > \frac{2}{(2l+1)}$$

Thus if $x \leqslant \frac{2}{2l+1}$, then $\left|\frac{1}{x}-l\right| \not< \frac{1}{2}$. This is an awkward situation because our assertion is that if $\lim\limits_{x \to 0}\left(\frac{1}{x}\right)=l$, then for points near 0, $\left|\frac{1}{x}-l\right|$ has to be less than $\frac{1}{2}$. (In fact any number instead of $\frac{1}{2}$). But as seen above, for points such that $x \geqslant 2/(2l+1)$, $\left|\frac{1}{x}-l\right| \not< \frac{1}{2}$. Hence $\lim\limits_{x \to 0}\left(\frac{1}{x}\right) \neq l$. One can similarly discuss the cases $l<0$ and $l=0$. Thus the limit in question does not exist.

Definition 2·1. *Let f be a function defined on some open interval containing c, except possibly at c. f is said to approach (tend to) a limit l as x approaches (tends to) c if $f(x)$ can be made as close to l as we like by taking x sufficiently close to c.*

In symbols, we write $\lim\limits_{x \to c} f(x)=l$.

Remarks. 1. We have already observed that for $\lim\limits_{x \to c} f(x)$ to exist, it is not at all necessary that f be defined at $x=c$. It is enough that f is defined in some open interval containing c.

2. No function can tend to two differant limits as $x \to c$. In other words, if $\lim\limits_{x \to c} f(x)=l$, and $\lim\limits_{x \to c} f(x)=m$, then $l=m$.

2·3·1. Some important results on limits

We now state (mostly without proof) some important results on limits.

Let f and g be defined on some open interval I containing c, but not necessarily at $x=c$. Also let $\lim\limits_{x \to c} f(x)=l$, $\lim\limits_{x \to c} g(x)=m$, and let k be some fixed real number, Then

(*i*) $\lim\limits_{x \to c} (f+g)(x)$ exists and equals $l+m$,

i.e.,

$$\lim_{x \to c} [f(x)+g(x)]=\lim_{x \to c} f(x)+\lim_{x \to c} g(x).$$

(*ii*) $\lim\limits_{x \to c} (fg)(x)$ exists and equals lm, *i.e.*,

$$\lim_{x \to c} [f(x)\, g(x)]=[\lim_{x \to c} f(x)]\,[\lim_{x \to c} g(x)]$$

(*iii*) $\lim\limits_{x \to c} (kf)(x)$ exists and equals kl, *i.e.*,

$$\lim_{x\to c} (kf)(x)=k \lim_{x\to c} f(x)$$

(*iv*) **Quotient theorem for limits**

If $g(x)\neq 0$ anywhere on I and $m\neq 0$, then

$\lim_{x\to c} (f/g)(x)=l/m$, *i.e.*,

$\lim_{x\to c} [f(x)/g(x)]=\lim_{x\to c} f(x) / \lim_{x\to c} g(x)$, provided $\lim_{x\to c} g(x)\neq 0$.

(*v*) $\lim_{x\to c} [\,|f|\,(x)]=|\,l\,|$, *i.e.*,

$\lim_{x\to c} [\,|f|\,(x)]=\lim_{x\to c} |f(x)|$

Remark. By using the principle of mathematical induction, (*i*) and (*ii*) can be extended to the case of more than two functions, *i.e.*, we can show that if $f_1, f_2, \ldots\ldots\ldots, f_n$ be real functions defined on an open interval I containing c but not necessarily at $x=c$, then we have the following :

(*iv*) **Sum theorem for limits.** $\lim_{x\to c} [f_1(x)+f_2(x) \ldots\ldots\ldots +f_n(x)]$

$=\lim_{x\to c} f_1(x)+\lim_{x\to c} f_2(x)+\ldots\ldots\ldots \lim_{x\to c} f_n(x)$,

(*vii*) **Product theorem for limits.** $\lim_{x\to c} [f_1(x) f_2(x) \ldots\ldots f_n(x)]$

$=[\lim_{x\to c} f_1(x)]\ [\lim_{x\to c} f_2(x)] \ldots\ldots\ldots [\lim_{x\to c} f_n(x)]$, (A)

provided that $\lim_{x\to c} f_1(x)$, $\lim_{x\to c} f_2(x), \ldots\ldots\ldots \lim_{x\to c} f_n(x)$ all exist. (Roughly speaking, the above results assest that under certain conditions limit of the sum is equal to the sum of limits, and limit of the product is equal to the product of limits.)

Limits of polynomials

Let us consider the identity function

$f: x\to x$ for all $x\in \mathbf{R}$.

Letting $f_1=f_2=\ldots\ldots\ldots\ldots=f_n=f$, by A above,

$\lim_{x\to c} [x.x \ldots\ldots\ldots x, n \text{ times}]=(\lim_{x\to c} x)^n$, *i.e.*,

$$\boxed{\lim_{x\to c} x^n = c^n}$$

(*viii*) If a_n be a fixed real number, then by using (*iii*) above, we have

$$\lim_{x\to c} (a_n x^n) = a_n \lim_{x\to c} x^n = a_n c^n$$

i.e.,

$$\boxed{\lim_{x\to c} (a_n x^n) = a_n c^n}$$

We are now ready to state (and prove) an important limit theorem for polynomial functions.

Limit theorem for polynomials. If $a_0, a_1, \ldots\ldots, a_n$ be any real numbers, then

(*ix*)
$$\boxed{\begin{aligned}&\lim_{x\to c} (a_n x^n + a_{n-1} x^{n-1} + \ldots\ldots + a_0)\\ &\qquad = a_n c^n + a_{n-1} c^{n-1} + \ldots\ldots + a_0\end{aligned}}$$

To prove the above result we have only to observe that by the 'Sum theorem for limits',

$$\lim_{x\to c} (a_n x^n + a_{n-1} x^{n-1} + \ldots\ldots + a_0)$$

$$= \lim_{x\to c} (a_n x^n) + \lim_{x\to c} (a_{n-1} x^{n-1}) + \ldots\ldots\ldots + \lim_{x\to c} (a_0)$$

$$= a_n c^n + a_{n-1} c^{n-1} + \ldots\ldots + a_0$$

The result established in (*ix*) above is rather important. It tells us that for each real number c, the limit of every polynomial function f exists at $x=c$, and furthermore, it says that the limit is precisely $f(c)$.

Limit of rational functions

Let us consider the rational function r defined by

$$r(x) = \frac{f(x)}{g(x)},$$

where $f(x)$ and $g(x)$ are the polynomials

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots\ldots\ldots + a_0,$$

$$g(x) = b_m x^m + b_{m-1} x^{m-1} + \ldots\ldots\ldots + b_0.$$

By the limit theorem for polynomials,

$$\lim_{x\to c} f(x) = f(c), \text{ and } \lim_{x\to c} g(x) = g(c)$$

Suppose that $g(c) \neq 0$. Then, by the 'quotient theorem for limits',

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \frac{f(c)}{g(c)}, \text{ i.e.,}$$

$$\boxed{\lim_{x \to c} \frac{a_n x^n + a_{n-1} x^{n-1} + \ldots\ldots\ldots\ldots + a_0}{b_m x^m + b_{m-1} x^{m-1} + \ldots\ldots\ldots\ldots + b_0} = \frac{a_n c^n + a_{n-1} c^{n-1} + \ldots\ldots\ldots + a_0}{b_m c^m + b_{m-1} c^{m-1} + \ldots\ldots\ldots + b_0}}$$

provided $b_m c^m + b_{m-1} c^{m-1} + \ldots\ldots + b_0 \neq 0$.

In case $b_m c^m + b_{m-1} c^{m-1} \ldots\ldots + b_0 = 0$, the above result cannot be applied. In that case we shall have to proceed in a sightly different way. The following examples will illustrate the method. Before we take up the examples, we give below two important results which are often found useful.

Sandwich theorem of limits.

Let f, g, and h be three real functions defined at all points of an open interval I containing a point c, but not necessarily at $x=c$, and let $f(x) < h(x) \leqslant g(x)$ for all $x \in I \sim \{c\}$.

If $\lim_{x \to c} f(x)$ and $\lim_{x \to c} g(x)$ both exist, then $\lim_{x \to c} h(x)$ exists, and

(*xi*)

$$\boxed{\lim_{x \to c} f(x) < \lim_{x \to c} h(x) < \lim_{x \to c} g(x)}$$

(*xiii*) **A fundamental limit theorem**

Theorem 2·2. *If n be any natural number, then*

$$\lim_{x \to c} \frac{x^n - c^n}{x - c} = nc^{n-1} \ (c > 0)$$

Proof. Four different cases arise :

Case 1. n=0. The equality holds since each sides equals 0.

Case 2. ***n* is a positive integer.**

By actual division,

$$\frac{x^n - c^n}{x - c} = x^{n-1} + x^{n-2}c + x^{n-3}c^2 + \ldots\ldots + c^{n-1} \qquad \ldots(1)$$

The process of division is valid so long as $x - c \neq 0$, *i.e.*, so long as $x \neq c$. Since in order to calculate the limit as $x \to c$ we have to consider values of x close to c but not the value c, therefore we can use (1) for calculating the limit of the left hand side as $x \to c$.

Therefore $\lim\limits_{x\to c}\dfrac{x^n-c^n}{x-c}=\lim\limits_{x\to c}(x^{n-1}+x^{n-2}c+\dots\dots+c^{n-1})$

$$=nc^{n-1},$$

since the limit of each of n terms x^{n-1}, x^{n-2} $c,\dots\dots, c^{n-1}$ as $x\to c$ is c^{n-1}.

Case 3. Let $n=p/q$, where p and q are positive integers. Put $x=y^q$, $c=b^q$ so that when $x=c$, then $y=b$.

Now $$\frac{x^n-c^n}{x-c}=\frac{y^{nq}-b^{nq}}{y^q-b^q}=\frac{y^p-b^p}{y^q-b^q}=\left(\frac{y^p-b^p}{y-b}\right)\Big/\left(\frac{y^q-b^q}{y-b}\right) \quad ...(2)$$

Now $$\lim_{y\to b}\frac{y^p-b^p}{y-b}=pb^{p-1},$$

$$\lim_{y\to b}\frac{y^q-b^q}{y-b}=qb^{q-1},$$

and if $x\to c$, then $y\to b$,

therefore by using (2), we have

$$\lim_{x\to b}\frac{x^n-c^n}{x-c}=\frac{pb^{p-1}}{qb^{q-1}}=\frac{p}{q}b^{p-q}=\frac{p}{q}(c^{1/q})^{p-q}=nc^{n-1},$$

since $p/q=n$.

Therefore, in this case also,

$$\lim_{x\to c}\frac{x^n-c^n}{x-c}=nc^{n-1}.$$

Case 4. Let $n<0$. Put $n=-m$, so that m is either a positive integer or a positive fraction.

$$\therefore \qquad \frac{x^n-c^n}{x-c}=\frac{x^{-m}-c^{-m}}{x-c}$$

$$=-\frac{1}{x^mc^m}\cdot\frac{x^m-c^m}{x-c},$$

so that $$\lim_{x\to c}\frac{x^n-c^n}{x-c}=\lim_{x\to c}\left(-\frac{1}{x^mc^m}\cdot\frac{x^m-c^m}{x-c}\right)$$

$$=\lim_{x\to c}\left(-\frac{1}{x^mc^m}\right)\lim_{x\to c}\left(\frac{x^m-c^m}{x-c}\right)$$

$$=-\frac{1}{c^{2m}}.mc^{m-1}, \text{ by case (1) and (2)}$$

$$=(-m)c^{-m-1}$$

$$=nc^{n-1}, \text{ since } n=-m.$$

Example 9. *Find* $\lim_{x\to 2} (x^3+4x^2-3x+7)$.

Solution.

$$\lim_{x\to 2} (x^3+4x^2-3x+7)=\lim_{x\to 2} x^3+\lim_{x\to 2} (4x^2)+\lim_{x\to 2} (-3x)+\lim_{x\to 2} (7)$$
$$=2^2+4.2^2+(-3).2+7$$
$$=25.$$

Aliter. If $f(x)=x^3+4x^2-3x+7$, then by the limit theorem for polynomials.

$$\lim_{x\to 2} f(x)=f(2)=2^3+4.2^2-3.2+7=25.$$

Example 10. Evaluate $\lim_{x\to 2} \{(x^2-2x+3)(x^2+3x+1)\}$.

Solution. By the product theorem for limits, we have

$$\lim_{x\to 2} \{(x^2-2x+3)(x^2+3x+1)\}$$
$$=\lim_{x\to 2} (x^2-2x+3) \lim_{x\to 2} (x^2+3x+1),$$
$=(2^2-2.2+3) . (2^2+3.2+1)$, by the limit theorem for polynomials,
$$=3.11=33.$$

Example 11. *Show that* $\lim_{x\to 0} \left(\frac{x^2+9}{x^2+3}\right)=3$.

Solution. Let $f(x)=x^2+9$, $g(x)=x^2+3$.

Now $\lim_{x\to 0} f(x)=9$, $\lim_{x\to 0} g(x)=3$.

Since $\lim_{x\to 0} g(x)\neq 0$, by the quotient theorem for limits,

$$\lim_{x\to 0} \frac{f(x)}{g(x)}=\lim_{x\to 0} f(x)/ \lim_{x\to 0} g(x)$$
$$=\frac{9}{3}=3.$$

Example 12. *Find* $\lim_{x\to 4} \frac{1/x-1/4}{x-4}$.

Solution. Let us define a function f by setting

$$f(x)=\frac{1/x-1/4}{x-4}, \ (x\neq 4).$$

We are required to find $\lim_{x\to 4} f(x)$.

Since f is not defind at $x=4$, therefore we cannot find the limit by substituting $x=4$ in $f(x)$. However, if $x \neq 4$, then

$$f(x)=\frac{4-x}{4x(x-4)}=\frac{-(x-4)}{4x(x-4)}=-\frac{1}{4x}, \text{ by cancelling}$$

out the non-zero factor $x-4$ from the numerator and the denominator.

Now $-1/(4x)$ is close to $-\frac{1}{16}$ if x is close to 4, and therefore $f(x)$ is close to $-\frac{1}{16}$ if x is close to 4, (but unequal to it). Therefore

$$\lim_{x \to 4} \frac{1/x-1/4}{x-4}=-\frac{1}{16}.$$

Example 13. *Find* $\lim_{h \to 0} \frac{(3+h)^3-27}{h}$.

Solution. If the function f be defined by

$$f(h)=\frac{(3+h)^3-27}{h} \qquad (h \neq 0),$$

then we are required to find

$$\lim_{h \to 0} f(h).$$

Since $f(h)$ is not defined for $h=0$, therefore we cannot find the limit by substituting $h=0$ in $f(h)$. Let us therefore proceed by simplifying $f(h)$.

$$\begin{aligned} f(h) &= \frac{(3+h)^3-27}{h}, \\ &= \frac{(27+27h+9h^2+h^3)-27}{h}, \\ &= \frac{27h+9h^2+h^3}{h}, \\ &= 27+9h+h^2, \text{ since } h \neq 0 \end{aligned}$$

Now $f(h)$ is close to 27 when h is close to 0 but unequal to it. Therefore $\lim_{h \to 0} f(h)=27$.

Example 14. *Find* $\lim_{x \to 1} \frac{\sqrt{x+4}-\sqrt{5}}{x-1}$.

Solution. Let us define a function f by setting

$$f(x)=\frac{\sqrt{x+4}-\sqrt{5}}{x-1} \qquad (x\neq 1)$$

Since f is not defined at $x=1$, we cannot find the limit by substituting $x=1$ in $f(x)$. We shall try to put $f(x)$ in a suitable form by *rationalizing the numerator*, *i.e.*, by multiplying the numerator and denominator of $f(x)$ by $\sqrt{x+4}+\sqrt{5}$. By doing so, we have

$$f(x)=\frac{(\sqrt{x+4}-\sqrt{5})(\sqrt{x+4}+\sqrt{5})}{(x-1)(\sqrt{(x+4)}+\sqrt{5})}$$

$$=\frac{(x-1)}{(x-1)(\sqrt{(x+4)}+\sqrt{5})}$$

$$=\frac{1}{\sqrt{(x+4)}+\sqrt{5}}.$$

Observe that we have cancelled $x-1$ from the numerator and the denominator because $x\neq 1$.

Now $f(x)$ can be seen to be close to $\frac{1}{2\sqrt{5}}$ when x is close to 1 (but not equal to 1). Therefore

$$\lim_{x\to 1}\frac{\sqrt{x+4}-\sqrt{5}}{x-1}=\frac{1}{2\sqrt{5}}.$$

EXERCISE 2 (*c*)

In each of the following problems, find the limit :

1. $\lim_{x\to 1}(2x+9)$.

2. $\lim_{x\to 2}\frac{x^2+1}{x+1}$.

3. $\lim_{x\to 0}(2x+3x+10)$.

4. $\lim_{x\to 1}\frac{x^2-x+1}{x^2+x+1}$.

5. $\lim_{x\to 4}\frac{x^2-16}{x-4}$.

6. $\lim_{x\to 2}\{(x^2+2)(x^3-3)\}$.

7. $\lim_{x\to 1}\frac{x^3-3x+2}{x^2-1}$.

8. $\lim_{x\to 2}\frac{x^3-6x^2+11x-6}{x^2-x-2}$.

9. $\lim_{x\to 3}\frac{\frac{1}{x+3}-\frac{1}{6}}{x-3}$.

10. $\lim_{x\to 9}\frac{\frac{1}{\sqrt{x}}-3}{x-9}$.

11. $\lim_{h\to 0}\frac{\sqrt{4+h}-2}{h}$.

12. $\lim_{x\to 1}\frac{\sqrt{x}-1}{x-1}$.

13. If $f(x)=x^2-3x+4$, find $\lim_{x\to 1}\frac{f(x)-f(1)}{x-1}$.

14. If $f(x)=\sqrt{25-x^2}$, find $\lim\limits_{x\to 3} \frac{f(x)-f(3)}{x-3}$.

15. Evaluate $\lim\limits_{h\to 1} \frac{h^3-1}{4-\sqrt{(h^2+15)}}$.

One-Sided Limits

While discussing the concept of the limit of a function as x tends to c, we do not bother as to whether the values taken by x are greater than c or less than c. We are only concerned with the fact that the values are arbitrarily close to c. By placing the additional restriction that the values are either all greater than c or all less than c we get the concept of a one-sided limit.

Meaning of $x\to c+$ *and* $x\to c-$

If a variable x takes values arbitrarily close to c but always greater than c, then we say that x tends to c from the right and write it as $x\to c+$. Similarly, if x takes values arbitrarily close to c but always less than c, then we say that x tends to c from the left and write it as $x\to c-$.

Meaning of $\lim\limits_{x\to c+} f(x)$ *and* $\lim\limits_{x\to c-} f(x)$

If $f(x)$ can be made as close to l as we please by taking values of x sufficiently close to c but greater than c, we say that $f(x)$ tends to l as x tends to c from the right. In symbols, we write

$$\lim_{x\to c+} f(x)=l.$$

Similarly, if $f(x)$ can be made as close to l as we please by taking values of x sufficiently close to c but less than c, we say that $f(x)$ tends to l as x tends to c from the left. In symbols, we write

$$\lim_{x\to c-} f(x)=l.$$

Remark. Limit from the left is also called left-hand limit, and limit from the right is also called right-hand limit.

The following theorem, whose obvious proof is omitted, gives the relationship between

$$\lim_{x\to c+} f(x),\quad \lim_{x\to c-} f(x),\quad \lim_{x\to c} f(x).$$

Theorem 2.3. *A function f tends to a limit l as $x\to c$ if and only if*

$$\lim_{x\to c+} f(x) = \lim_{x\to c-} f(x)=l.$$

The above theorem is sometimes used to show that the limit of a function f exists as x tends to c.

The following theorem gives the relation between limits from the left and limits from the right.

Theorem 2·4. If $\lim_{x \to 0+} f(-x)$ exists, then $\lim_{x \to 0-} f(x)$ also exists, and both the limits are equal.

Proof. Suppose $\lim_{x \to 0+} f(-x)$ exists and equals l.

$$\lim_{x \to 0-} f(x) = \lim_{\substack{h \to 0 \\ h>0}} f(-h)$$

$$= \lim_{x \to 0+} f(-x)$$

$$= l.$$

Example 15. *Let f be defined by setting*

$$f(x) = \begin{cases} 2x-1 & , \text{ if } x<0 \\ 0 & , \text{ if } x=0 \\ 3-2x, & , \text{ if } x>0. \end{cases}$$

Find the limits from the left and the right as x tends to 0. Does f tend to a limit as x tends to 0 ?

Solution. When $x<0, f(x)=2x-1$. To find $\lim_{x \to 0-} f(x)$, we put $x=-h$ and take limits as $h \to 0$, h remaining positive.

$$\therefore \lim_{x \to 0-} f(x) = \lim_{h \to 0} 2(-h)-1=-1.$$

When $x>0$, $f(x)=3-2x$. To find limit $\lim_{x \to 0+} f(x)$, we put $x=h$ and take limits as $h \to 0$, h remaining positive.

$$\therefore \lim_{\Lambda \to 0+} f(x) = \lim_{h \to 0} 3-2h = 3.$$

Since $f(0+) \neq f(0-)$, therefore $\lim_{x \to 0} f(x)$ does not exist.

EXERCISE 2 (*d*)

Evaluate :

1. $\lim_{x \to 1-} x^2+1.$
2. $\lim_{x \to 2+} x^2-3x+1.$
3. $\lim_{x \to 0+} | x |.$
4. $\lim_{x \to 0-} 1-3x^2.$
5. $\lim_{x \to 2-} 4-x^2.$
6. $\lim_{\pi \to 3+} x^3-27.$
7. $\lim_{x \to 4-} | x-4 |.$
8. $\lim_{x \to 2-} x^2-2.$
9. $\lim_{x \to 1-} x^2 | x |.$
10. $\lim_{x \to 5-} x^3 | x |.$

SOME IMPORTANT LIMITS INVOLVING TRIGONOMENTRIC FUNCTIONS

In the following theorem we give some important limits involving the sine and cosine functions. These limits will be found useful later on.

Theorem 2·5. (*a*) $\lim\limits_{x\to 0} \sin x=0$,

(*b*) $\lim\limits_{x\to c} \sin x=\sin c$,

(*c*) $\lim\limits_{x\to c} \cos x=\cos c$,

(*d*) $\lim\limits_{x\to 0} \cos x=1$,

(*e*) $\lim\limits_{x\to 0} \dfrac{\sin x}{x}=1$.

Proof. (*a*) Since we have to find the limit of $\sin x$ as x tends to zero, therefore we can take x to be small. Let us assume that $x>0$, and x is small. Consider a rectangular co-ordinate system XOY. With O as centre draw a circle of unit radius meeting OX at A. Let P be a point on the circle such that the radian measure of $\angle$POA is $2x$. Then A is the point $(1, 0)$ and P is the point $(\cos 2x, \sin 2x)$.

Now $AP=\sqrt{\{(\cos 2x-1)^2+\sin^2 2x\}}$

$=2 \sin x>0$.

Also, length of the arc $AP=2x$.

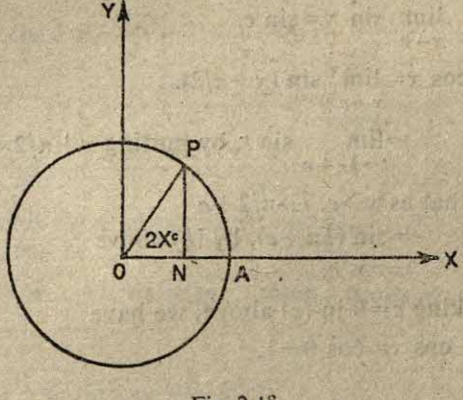


Fig. 2·18.

Since chord AP$<$ arc AP,

therefore $\quad 0<2 \sin x<2x$,

or $\quad 0<\sin x<x$.

Since the above inequality remains true however small x may be, therefore taking limits as $x \to 0+$, we have

$$0 \leqslant \lim_{x \to 0+} \sin x \leqslant \lim_{x \to 0+} x$$

or

$$0 \leqslant \lim_{x \to 0} \sin x \leqslant 0,$$

or

$$\lim_{x \to 0+} \sin x = 0. \qquad \ldots(1)$$

Let now $x < 0$. Putting $x = -t$, so that $t > 0$ and $t \to 0+$ as $x \to 0-$, we have

$$\lim_{x \to 0-} \sin x = \lim_{t \to 0+} \sin(-t) = \lim_{x \to 0+} [-\sin t] = 0, \qquad \ldots(2)$$

by (1).

From (1) and (2), we have

$$\lim_{x \to 0} \sin x = 0.$$

(*b*) If c be any real number, then

$$\sin x - \sin c = 2 \cos \tfrac{1}{2}(x+c) \sin \tfrac{1}{2}(x-c),$$

so that
$$0 \leqslant |\sin x - \sin c| = |2 \cos \tfrac{1}{2}(x+c) \sin \tfrac{1}{2}(x-c)|$$
$$\leqslant 2\,|\sin \tfrac{1}{2}(x-c)|.$$

Taking limits as $x \to c$, we have

$$0 \leqslant \lim_{x \to c} |\sin x - \sin c| \leqslant \lim_{x \to c} |\sin \tfrac{1}{2}(x-c)| = 0.$$

$$\therefore \quad \lim_{x \to c} |\sin x - \sin c| = 0,$$

or

$$\lim_{x \to c} \sin x = \sin c.$$

(*c*) $\lim\limits_{x \to c} \cos x = \lim\limits_{x \to c} \sin(x + \pi/2)$.

$$= \lim_{t \to \frac{1}{2}\pi + c} \sin t, \text{ by putting } x + \pi/2 = t$$

and observing that as $x \to c$, $t \to \pi/2 + c$

$$= \sin(\tfrac{1}{2}\pi + c), \text{ by } (b) \text{ above}$$
$$= \cos c.$$

(*d*) By taking $c = 0$ in (*c*) above, we have

$$\lim_{x \to 0} \cos x = \cos 0 = 1.$$

(*e*) Let us first suppose that $x > 0$. Since we have to take limits as $x \to 0$, we may assume that $0 < x < \pi/2$. Consider an $\angle$MOA whose radian measure is x. Let $\angle$MOA be placed in standard position relative to the axes of coordinates so that OA lies along OX,

and O is at the origin of co-ordinates. Let A and M be on the unit circle centred at O, so that OA=OM=1, length of the arc AM=x, and area of the sector AOM=$x/2$. Draw MP$\perp$OA and let the perpendicular from A on OX meet OM produced at N.

Since area of $\triangle$ OPM < area of sector OAM < area of $\triangle$OAN, therefore,

$$\tfrac{1}{2}\sin x\cos x < \tfrac{1}{2}x < \tfrac{1}{2}\tan x,$$

or $$\cos x < \frac{\sin x}{x} < \frac{1}{\cos x} \qquad ...(1)$$

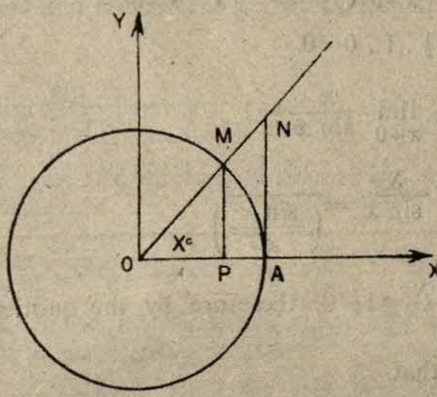


Fig. 2·19.

Inequalities (1) remain true however small x may be.

Let now $x<0$. Writing $x=-t$, and observing that $t>0$, we have from (1)

$$\cos t < \frac{\sin t}{t} < \frac{1}{\cos t}$$

or $$\cos(-x) < \frac{\sin(-x)}{-x} < \frac{1}{\cos(-x)},$$

or $$\cos x < \frac{\sin x}{x} < \frac{1}{\cos x}. \qquad ...(2)$$

Thus (1) holds even when $x<0$, provided $|x|<\pi/2$, howsoever small $|x|$ may be.

Since $\lim\limits_{x\to 0}\cos x=1$, therefore, taking limits as $x\to 0$, we have from (1) and (2),

$$1 \leqslant \lim_{x\to 0}\frac{\sin x}{x} \leqslant 1,$$

i.e. $$\lim_{x\to 0}\frac{\sin x}{x}=1.$$

Corollary 1. $\lim\limits_{x\to 0} \dfrac{1-\cos x}{x}=0.$

Proof. $\lim\limits_{x\to 0} \dfrac{1-\cos x}{x}=\lim\limits_{x\to 0} \dfrac{2\sin^2(\frac{1}{2}x)}{x}$

$$=\lim_{x\to 0}\left[\frac{1}{2}\left(\frac{\sin x/2}{x/2}\right)^2 . x\right],$$

$$=\tfrac{1}{2}\lim_{x\to 0}\left\{\frac{\sin (x/2)}{x/2}\right\}^2 . \lim_{x\to 0} x,$$

$$=\tfrac{1}{2} . 1 . 0=0.$$

Corollary 2. $\lim\limits_{x\to 0} \dfrac{x}{\sin x}=1$

Proof. $$\frac{x}{\sin x}=\frac{1}{\left(\dfrac{\sin x}{x}\right)}$$

Since $\lim\limits_{x\to 0} \dfrac{\sin x}{x}=1\neq 0$, therefore by the quotient theorem for limits it follows that

$$\lim_{x\to 0} \frac{x}{\sin x}=\frac{\lim\limits_{x\to 0} 1}{\lim\limits_{x\to 0} (\sin x/x)}=\frac{1}{1}=1.$$

Corollary 3. $\lim\limits_{x\to 0} \dfrac{\tan x}{x}=1.$

Proof. Writing $\dfrac{\tan x}{x}=\dfrac{\sin x}{x} . \dfrac{1}{\cos x}$, we have

$$\lim_{x\to 0} \frac{\tan x}{x}=\lim_{x\to 0} \frac{\sin x}{x} . \lim_{x\to 0}\left(\frac{1}{\cos x}\right)$$

$$=\lim_{x\to 0} \frac{1}{\cos x}, \text{ since } \lim_{x\to 0} \frac{\sin x}{x}=1$$

$$=1, \text{ since } \lim_{x\to 0} \cos x=1.$$

Example 16. *Evaluate* $\lim\limits_{x\to 0} \dfrac{\sin 5x}{\sin 3x}$

Solution. $\dfrac{\sin 5x}{\sin 3x}$

$$=\frac{5\left(\dfrac{\sin 5x}{5x}\right)}{3\left(\dfrac{\sin 3x}{3x}\right)} \quad \text{...(1)}$$

Now $\lim\limits_{x\to 0} \dfrac{\sin 5x}{5x}=\lim\limits_{t\to 0} \dfrac{\sin t}{t}=1,$ ...(2)

by putting $5x=t$ and noting that if $x\to 0$ then $t\to 0$

Similarly

$$\lim_{x\to 0} \frac{\sin 3x}{3x}=1 \quad \text{...(3)}$$

Using (2) and (3), we have from (1), by using the quotient theorem for limits

$$\lim_{x\to 0} \frac{\sin 5x}{\sin 3x}=\frac{5.1}{3.1}=\frac{5}{3}$$

Remark. After a little practice you will see that the above working can be re-arranged as follows :

$$\lim_{x\to 0} \frac{\sin 5x}{\sin 3x}=\lim_{x\to 0}\left(\frac{5}{3}\cdot\frac{\sin 5x}{5x}\cdot\frac{3x}{\sin 3x}\right)$$

$$=\left(\lim_{x\to 0} \frac{\sin 5x}{5x}\right)\left(\lim_{x\to 0} \frac{3x}{\sin 3x}\right)$$

$$=\frac{5}{3}.\ 1.1=\frac{5}{3}.$$

Example 17. *Evaluate* $\lim\limits_{x\to 0} \dfrac{\sin x}{x-\pi}$. (*A.I.S.S.C.E. 1985*)

Solution. By putting $x-\pi=t$ and observing that when $x\to\pi$, $t\to 0$, we find that

$$\lim_{x\to 0} \frac{\sin x}{x-\pi}=\lim_{t\to 0} \frac{\sin(\pi+t)}{t}$$

$$=\lim_{t\to 0}\left(-\frac{\sin t}{t}\right)$$

$$=-\lim_{t\to 0} \frac{\sin t}{t}$$

$$=-1$$

EXERCISE 2 (*e*)

Evaluate :

1. $\lim\limits_{x\to 0} \dfrac{\sin 2x}{3x}$.   2. $\lim\limits_{x\to 0} \dfrac{5x}{\sin 3x}$.

3. $\lim\limits_{x\to 0} \dfrac{\tan 4x}{5x}$.   4. $\lim\limits_{x\to 0} \dfrac{3x}{\tan 2x}$.

5. $\lim\limits_{x\to 0} \dfrac{\sin 5x}{\sin 4x}$.   6. $\lim\limits_{x\to 0} \dfrac{\tan 3x}{\tan 4x}$.

7. $\lim\limits_{x\to 0} \dfrac{\sin 5x-\sin 3x}{\sin x}$. (*D.B.S.S C.E. 1984*)

8. $\lim\limits_{x\to 0} \dfrac{x^3}{\sin (x^2)}$. (*A.I.S.S.C.E. 1984*)

9. $\lim\limits_{x\to 0} \dfrac{\sin 2x\,(1-\cos 2x)}{x^3}$. (*A.I.S.S C.E. 1986*)

10. $\lim\limits_{x\to 0} \dfrac{1-\cos x}{\sin^2 x}$. (*D.B.S.S.C.E. 1987*)

11. $\lim\limits_{x\to 0} \dfrac{1-\cos x}{x}$. (*A.I.S S.C.E. 1987*)

12. $\lim\limits_{x\to 0} \dfrac{x^3 \cot x}{1-\cos x}$. (*D B.S.S.C.E. 1988*)

13. $\lim\limits_{x\to 0} \dfrac{\sqrt{1+x}-\sqrt{1-x}}{\sin^{-1} x}$. (*D.B.S.S.C E. 1989*)

2·6. SOME IMPORTANT LIMITS INVOLVING EXPONENTIAL AND LOGARITHMIC FUNCTIONS

In the following theorems we give three important limits involving the exponential and logarithmic functions. These limits will be found useful later on.

Theorem 2·6. $\lim\limits_{x\to 0} e^x=1$

Proof.

Step 1. We shall show that if $0 \leqslant x \leqslant 1$, then $1 \leqslant e^x \leqslant 1+ex$.

Since $\qquad e^x=1+x+\dfrac{x^2}{2\,!}+\dfrac{x^3}{3\,!}+\dots\dots,$

therefore if $x>0$, then each term of the above series is positive, and consequently $e^x > 1$.

Also, when $x=0$, $e^x=1$.

Thus $x \geqslant 0 \Rightarrow e^x \geqslant 1$ ...(1)

Also $$e^x=1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots\ldots$$

$$\leqslant 1+\frac{x}{1!}+\frac{x}{2!}+\frac{x}{3!}+\ldots\ldots$$

($\because$ $x^n \leqslant x$ whenever $0 \leqslant x \leqslant 1$)

$$=1+x\left(\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}+\ldots\ldots\right)$$

$$=1+x\,(e-1)$$

$$\leqslant 1+x\,e,\text{ since } x \geqslant 0 \qquad \ldots(2)$$

From (1) and (2), we find that

$$1 \leqslant e^x \leqslant 1+ex. \qquad \ldots(3)$$

Step 2. Taking limits as $x \to 0+$, we find from (3),

$$1 \leqslant \lim_{x \to 0+} e^x \leqslant 1,$$

or

$$\lim_{x \to 0+} e^x=1. \qquad \ldots(4)$$

Step 3. To find $\lim\limits_{x \to 0-} e^x$, let us observe that $\lim\limits_{x\to 0-} f(x)=\lim\limits_{x\to 0+} f(-x)$, and apply it to the function defined by $f(x)=e^x$.

$$\lim_{x\to 0-} e^x=\lim_{x\to 0+} e^{-x}=\lim_{x\to 0+}\frac{1}{e^x}=1 \qquad \ldots(5)$$

Step 4. From (4) and (5), we have

$$\lim_{x\to 0+} e^x=\lim_{x\to 0-} e^x=1,$$

so that

$$\lim_{x\to 0} e^x=1.$$

Theorem 2·7. $\lim\limits_{x\to 0}\dfrac{e^x-1}{x}=1.$

Proof.

Step 1. We shall show that if $0 \leqslant x \leqslant 1$,

then $1+x \leqslant e^x \leqslant 1+x+x^2\,(e-2)$

Since $x \geqslant 0$, therefore every term of the series expression for e^x is non-negative. Keeping the first two terms and leaving out the remaining terms, we have

$$e^x \geqslant 1+x \qquad \ldots(1)$$

Also,

$$e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\ldots\ldots$$

$$\leqslant 1+x+\frac{x^2}{2!}+\frac{x^2}{3!}+\frac{x^2}{4!}+\cdots,$$

(since $x^2 \geqslant x^n$ whenever $n \geqslant 2$)

$$=1+x+x^2\,(e-2). \qquad \ldots(2)$$

From (1) and (2), we find that

$$1+x \leqslant e^x \leqslant 1+x+x^2\,(e-2). \qquad \ldots(3)$$

Step 2. From (3) above, we have

$$1 \leqslant \frac{e^x-1}{x} \leqslant 1+x\,(e-2) \qquad \ldots(4)$$

Taking limits as $x \to 0+$, we have from (4),

$$1 \leqslant \lim_{x\to 0+} \frac{e^x-1}{x} \leqslant 1,$$

so that

$$\lim_{x\to 0+} \frac{e^x-1}{x}=1. \qquad \ldots(5)$$

Step 3. Let us write $f(x)=\dfrac{e^x-1}{x}$.

Then $f(-x)=\dfrac{e^{-x}-1}{-x}=\dfrac{1-e^x}{-x\,e^x}=\dfrac{e^x-1}{x}\cdot\dfrac{1}{e^x}$.

Taking limits as $x \to 0+$, we have

$$\lim_{x\to 0+} f(-x)=\lim_{x\to 0+}\frac{e^x-1}{x}\cdot\lim_{x\to 0+}\frac{1}{e^x}$$

$=1\,.\,1$, by step (2) above and Theorem 2·6

$=1$,

so that $$\lim_{x\to 0-} f(x)=\lim_{x\to 0+} f(-x)=1. \qquad \ldots(6)$$

From (5) and (6), we have

$$\lim_{x\to 0+} f(x)=\lim_{x\to 0-} f(x)=1,$$

i.e., $$\lim_{x\to 0} f(x)=1.$$

Thus $$\lim_{x\to 0}\frac{e^x-1}{x}=1.$$

Corollary. $\lim\limits_{x\to 0} \dfrac{x}{e^x-1}=1.$

Theorem 2·8. $\lim\limits_{x\to 0} \dfrac{\ln(1+x)}{x}=1.$

Proof.

Step 1. We shall first show that

$$\lim_{x\to 0} \ln(1+x)=0.$$

Since $1 < 1+x < e^x$ for all $x > 0$, ...(1)

therefore $\ln 1 < \ln(1+x) < x$, for all $x > 0$

i.e., $0 < \ln(1+x) < x$, for all $x > 0$.

so that $0 \leqslant \lim\limits_{x\to 0+} \ln(1+x) \leqslant 0$,

i.e.,

$$\lim_{x\to 0+} \ln(1+x)=0. \quad ...(2)$$

Step 2. We shall show that

$$\lim_{x\to 0} \ln(1+x)=0.$$

Proceeding in the same manner as above, we can show that if $-1 < x < 0$, then

$$e^x < 1+x < 1,$$

so that $x < \ln(1+x) < 0$

Taking limits as $x\to 0-$, we have

$$0 \leqslant \lim_{x\to 0-} \ln(1+x) \leqslant 0 \quad ...(3)$$

From (2) and (3), we have

$$\lim_{x\to 0-} \ln(1+x)=0. \quad ...(4)$$

Step 3. From (2) and (4) above, we have

$$\lim_{x\to 0+} \ln(1+x)=\lim_{x\to 0-} \ln(1+x)=0,$$

so that

$$\lim_{x\to 0} \ln(1+x)=0. \quad ...(5)$$

Step 4. Put $t=\ln(1+x)$.

Then $\dfrac{\ln(1+x)}{x}=\dfrac{t}{e^t-1}$ ...(6)

As $x\to 0$, $t\to 0$, by step 3.

$$\therefore \quad \lim_{x\to 0} \frac{\ln(1+x)}{x}=\lim_{t\to 0} \frac{t}{e^t-1}=1.$$

EXERCISE 2 (f)

Evaluate :

1. $\lim\limits_{x\to 0} e^{-2x}$.
2. $\lim\limits_{x\to 1} e^{x-1}$.
3. $\lim\limits_{x\to -1} e^{2x+1}$.
4. $\lim\limits_{x\to 2} e^{x^2-1}$.
5. $\lim\limits_{x\to 0} \dfrac{e^{2x}-1}{3x}$.
6. $\lim\limits_{x\to 0} \dfrac{4x}{e^{5x}-1}$.
7. $\lim\limits_{x\to 0} \ln(1-3x)$.
8. $\lim\limits_{x\to \frac{1}{3}} \ln(2-3x)$.
9. $\lim\limits_{x\to 0} \dfrac{\ln(1+2x)}{8x}$.
10. $\lim\limits_{x\to 0} \dfrac{5x}{\ln(1-3x)}$.

2·7. INFINITE LIMITS

Let f be a real function defined on a domain D containing the open interval $]\,c, d\,[$. We say that $f(x) \to +\infty$ *as* $x \to c+$ if $f(x)$ can be made as large as we please by taking x sufficiently close to c (but $> c$). In symbols, we then write

$$\lim_{\substack{x\to c\\ x>c}} f(x) = +\infty$$

Also, we say that $f(x) \to -\infty$ *as* $x \to c+$ *if* $f(x)$ *can be made as small as we please by taking* x *sufficiently close to* c (*but* $> c$). In symbols we write $\lim\limits_{\substack{x\to c\\ x>c}} f(x) = -\infty$.

Infinite limits from the left can be defined similarly. In fact, with the same notation as above we say that $f(x) \to +\infty$ *as* $x \to d-$ *if* $f(x)$ *can be made as large as we please by taking* x *sufficiently close to* d (*but* $< d$). In symbols, we then write

$$\lim_{\substack{x\to d\\ x<d}} f(x) = +\infty$$

Also, *we say that* $f(x) \to -\infty$ *as* $x \to d-$ *if* $f(x)$ *can be made as small as we please by taking* x *sufficiently close to* d (but $< d$). Our definitions of the above concepts are all intuitive. The following examples will illustrate the definitions.

Illustrations

1. Let f be defined by setting

$$f(x) = \frac{1}{x^2-1}, \text{ for all } x \in \mathbf{R} - \{-1, 1\}.$$

Case 1. Let $x > 1$. Then $x^2-1 > 0$. By taking x sufficiently close to 1, we can make x^2-1 as small as we like, which means that we can make $\frac{1}{x^2-1}$ as large as we like.

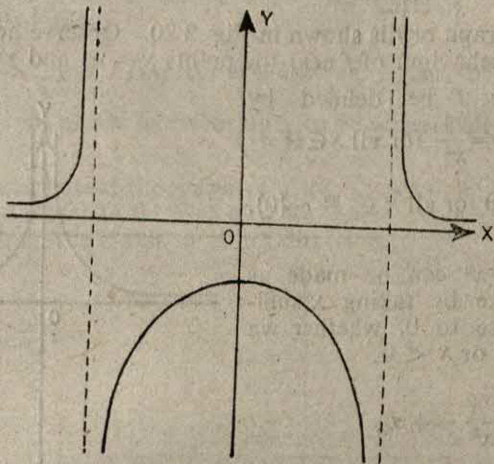


Fig. 2·20. Graph of $y=\frac{1}{x^2-1}$.

$\therefore \quad \lim_{x\to 1+} f(x)=+\infty$

Case 2. Let $-1 < x < 1$. Clearly $x^2-1 < 0$ and therefore $\frac{1}{x^2-1} < 0$.

(*a*) By taking x sufficiently close to 1, we can make $|x^2-1|$ as small as we like, *i.e.*, $\left|\frac{1}{x^2-1}\right|$ can be made as large as we like, so that $\frac{1}{x^2-1}$ can be made as small as we like.

$\therefore \quad \lim_{x\to 1-} f(x)=-\infty.$

(*b*) By taking x sufficiently close to -1, we can make $|x^2-1|$ as small as we like, *i.e.*, $\left|\frac{1}{x^2-1}\right|$ can be made as large as we like, *i.e.*, $\frac{1}{x^2-1}$ can be made as small as we like.

$\therefore \quad \lim_{x\to -1+} f(x)=-\infty$

Case 3. Let $x < -1$. Clearly $x^2-1 > 0$, By taking x sufficiently close to -1 we can make x^2-1 as close to 0 as we like, so that $\frac{1}{x^2-1}$ can be made as large as we like.

$$\therefore \quad \lim_{x \to -1-} f(x) = +\infty$$

The graph of f is shown in Fig. 2·20. Observe how the graph shows the behaviour of f near the points $x=-1$ and $x=+1$.

2. Let f be defined by setting $f(x)=\frac{1}{x^2}$ for all $x \in \mathbf{R} \sim \{0\}$. $\frac{1}{x^2} > 0$ for all $x \in \mathbf{R} \sim \{0\}$.

Since x^2 can be made as small we like by taking x sufficiently close to 0, whether we take $x > 0$ or $x < 0$, therefore

$$\lim_{x \to 0+} \frac{1}{x^2} = +\infty,$$

and $$\lim_{x \to 0-} \frac{1}{x^2} = +\infty$$

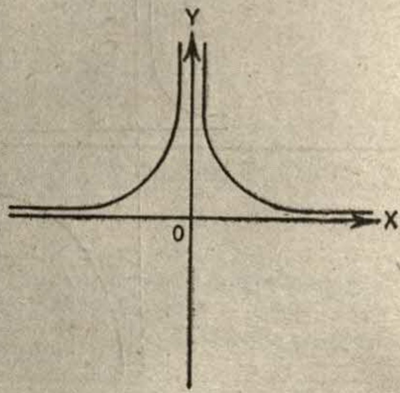


Fig. 2·21. Graph of $y=1/x^2$.

3. Let $f(x) = -\frac{1}{x^2}$ whenever $x \neq 0$.

It can be easily seen that (how ?)

$$\lim_{x \to 0+} f(x) = -\infty,$$

and $$\lim_{x \to 0-} f(x) = -\infty.$$

2·8. LIMITS AS $x \to +\infty$ (or $-\infty$)

(*i*) Let f be a real function defined on a domain D containing the open ray $]\, c, \infty\, [$. We say that *$f(x) \to l$ as $x \to +\infty$ if $f(x)$ can be made as close to l as we like by taking x sufficiently large*. In symbols, we write

$$\lim_{x \to +\infty} f(x) = l.$$

(*ii*) Let f be a real function defined on a domain D containing the open ray $]-\infty, c\,[$. We say that *$f(x) \to l$ as $x \to -\infty$ if $f(x)$ can be made as close to l as we like by taking x sufficiently large in absolute value but negative.* In symbols, we write

$$\lim_{x \to -\infty} f(x) = l.$$

Illustrations 1. Let $f(x)=\frac{1}{x^2+1}$ for all $x \in \mathbf{R}$.

Here $\lim\limits_{x \to +\infty} f(x)=0$, because $\frac{1}{x^2+1}$ can be made as close to zero as we please (by making x^2+1 very large) by choosing x to be sufficiently large.

Also, $\lim\limits_{x \to -\infty} f(x)=0$, because $\frac{1}{x^2+1}$ can be made as close to zero as we please by choosing x to be numerically very large but negative.

[Observe that the graph of f, as shown in Fig. 2·22, becomes closer and closer to the x-axis as we move further away from the origin, along the x-axis in either direction.]

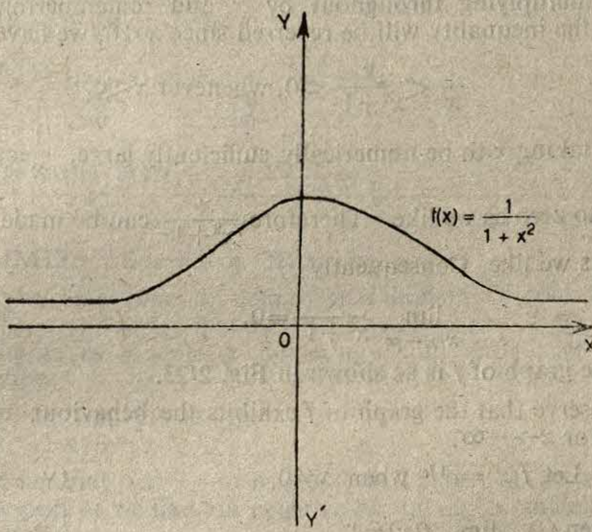


Fig. 2·22

2. Let $f(x)=\frac{x}{x^2+1}$ for all $x \in \mathbf{R}$.

It can be easily seen that

$$\operatorname*{Lim}_{x \to +\infty} f(x)=0=\lim_{x \to -\infty} f(x).$$

In fact, first let us take $x>0$.

Then $\qquad 0<\frac{x}{x^2+1}<\frac{1}{x}$

By taking x sufficiently large we can make $\frac{x}{x^2+1}$ as small as

we like. Since $f(x)$ can be made as close to zero as we like by choosing x sufficiently large, therefore it follows that

$$\lim_{x\to+\infty} f(x)=0.$$

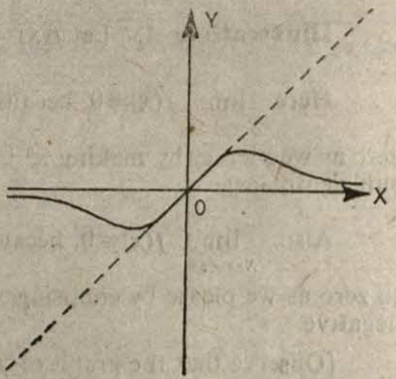


Fig. 2·23

Let us now take $x<0$. Since $x^2+1>x^2$ for all $x\in\mathbf{R}$, therefore

$$0<\frac{1}{x^2+1}<\frac{1}{x^2},$$

whenever $x\neq 0$.

By multiplying throughout by x and remembering that by doing so the inequality will be reversed since $x<0$, we have

$$\frac{1}{x}<\frac{x}{x^2+1}<0, \text{ whenever } x<0.$$

By taking x to be numerically sufficiently large, $\frac{1}{x}$ can be made as close to zero as we like. Therefore $\frac{x}{x^2+1}$ can be made as close to zero as we like. Consequently

$$\lim_{x\to-\infty} \frac{x}{x^2+1}=0.$$

The graph of f is as shown in Fig. 2·23.

Observe that the graph of f exhibits the behaviour of f when $x\to+\infty$ or $x\to-\infty$.

3. Let $f(x)=e^{1/x}$ when $x\neq 0$, $f(0)=0$.

Then (*i*) $\lim\limits_{x\to+\infty} f(x)=1.$

For, letting $t=\frac{1}{x}$, and observing that $x\to+\infty$ if and only if $t\to 0$ through positive values, we have

$$\operatorname{Lim}_{x\to+\infty} f(x)=\lim_{\substack{t\to 0\\ t>0}} f(1/t)$$
$$=\lim_{\substack{t\to 0\\ t>0}} e^t=1$$

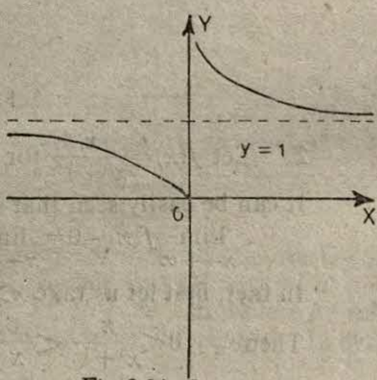


Fig. 2·24. Graph of $y=e^{1/x}$

(*ii*) $\lim_{x \to -\infty} f(x) = 1$.

For, letting $t = \frac{1}{x}$, and observing that $x \to -\infty$ if and only if $t \to 0$ through negative values, we have

$$\lim_{x \to -\infty} f(x) = \lim_{\substack{t \to 0 \\ t < 0}} f(1/t)$$
$$= \lim_{\substack{t \to 0 \\ t > 0}} e^t = 1.$$

4. Let $f(x) = x \sin \frac{1}{x}$, whenever $x \neq 0$.

Then $\lim_{x \to +\infty} f(x) = \lim_{x \to -\infty} f(x) = 1$.

For, as in illustration 3 above,

$$\lim_{x \to +\infty} f(x) = \lim_{\substack{t \to 0 \\ t > 0}} f(1/t) = \lim_{\substack{t \to 0 \\ t > 0}} \frac{\sin t}{t} = 1,$$

$$\lim_{x \to -\infty} f(x) = \lim_{\substack{t \to 0 \\ t < 0}} f(1/t) = \lim_{\substack{t \to 0 \\ t < 0}} \frac{\sin t}{t} = 1.$$

INFINITE LIMITS AS $x \to +\infty$ OR AS $x \to -\infty$

(*i*) Let f be a real function defined on a domain D containing the open ray $]c, \infty[$. We say that $f(x) \to +\infty$ as $x \to +\infty$ if $f(x)$ can be made as large as we please by taking x sufficiently large. In symbols, we write

$$\lim_{x \to +\infty} f(x) = +\infty$$

Also, we say that $f(x) \to -\infty$ as $x \to +\infty$ if $f(x)$ can be made numerically as large as we like but negative by taking x sufficiently large. In symbols, we write

$$\lim_{x \to +\infty} f(x) = -\infty$$

(*ii*) Let f be a real function defined on a domain D containing the open ray $]-\infty, c[$. We say that $f(x) \to +\infty$ as $x \to -\infty$ if $f(x)$ can be made as large as we please by taking x numerically sufficiently large but negative. In symbols, we write

$$\lim_{x \to -\infty} f(x) = +\infty$$

Also, we say that $f(x) \to -\infty$ as $x \to -\infty$ if $f(x)$ can be made numerically as large as we please but negative by taking x sufficiently large but negative.

The following illustrations will make the meaning of the above definitions clear.

Illustrations 1. Let $f(x)=x^2$ for all $x \in \mathbf{R}$.

Then $\lim_{x \to +\infty} f(x) = \lim_{x \to -\infty} f(x) = +\infty$.

The graph of f is as shown in Fig. 2·25.

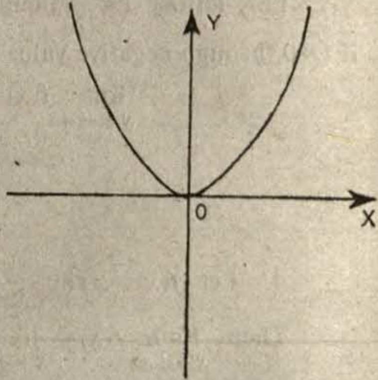


Fig. 2·25. Graph of $y=x^2$

2. Let $f(x)=-x^2$ for all $x \in \mathbf{R}$.

Then $\lim_{x \to +\infty} f(x) = \lim_{x \to -\infty} f(x) = -\infty$.

The graph of f is as shown in Fig. 2·26.

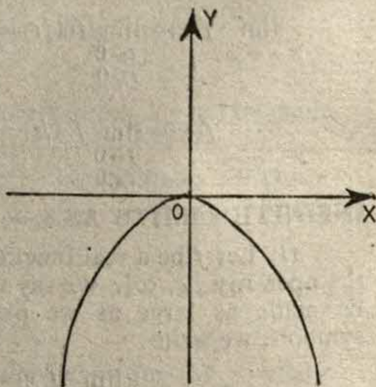


Fig. 2·26. Graph of $y=-x^2$

3. Let $f(x)=x^3$ for all $x \in \mathbf{R}$.

Then $\lim_{x \to +\infty} f(x) = +\infty$,

$\lim_{x \to -\infty} f(x) = -\infty$.

The graph of f is as shown in Fig. 2·27.

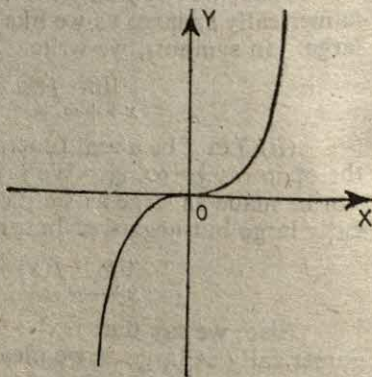


Fig. 2·27. Graph of $y=x^3$

4. Let $f(x)=-x^3$ for all $x \in \mathbf{R}$.

Then $\lim\limits_{x \to +\infty} f(x) = -\infty$,

$\lim\limits_{x \to -\infty} f(x) = +\infty$.

The graph of f is as shown in Fig. 2·28.

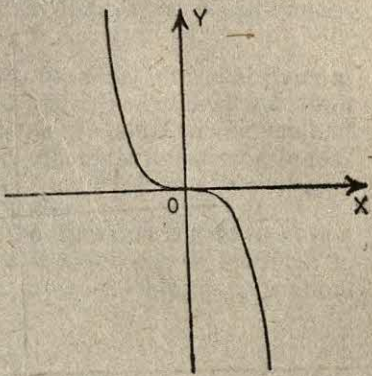


Fig. 2·28. Graph of $y=-x^3$.

EXERCISE 2 (g)

Evaluate :

1. $\lim\limits_{x \to 3+} \dfrac{1}{x^2-9}$

2. $\lim\limits_{x \to 4-} \dfrac{1}{x^2-16}$.

3. $\lim\limits_{x \to 0+} \dfrac{2}{2x^2+1}$

4. $\lim\limits_{x \to 2-} \dfrac{1}{4-x^2}$.

5. $\lim\limits_{x \to +\infty} \dfrac{1}{1-x^3}$.

6. $\lim\limits_{x \to -\infty} \dfrac{2}{2x^3+1}$.

7. $\lim\limits_{x \to \pi-} \cot x$.

8. $\lim\limits_{x \to -\infty} x^3+x^2+1$.

9. $\lim\limits_{x \to +\infty} -2x^4+1$.

10. $\lim\limits_{x \to -\infty} 5-x^3$.

2·9. CONTINUITY

Consider the real functions f and g be defined as follows :

$$f(x)=\begin{cases} x, & \text{if } 0 \leqslant x < 1. \\ 1, & \text{if } x \geqslant 1 \end{cases}$$

$$g(x)=\begin{cases} x, & \text{if } 0 \leqslant x < 1. \\ -1, & \text{if } x \geqslant 1 \end{cases}$$

The graphs of f and g are shown in Fig. 2·29.

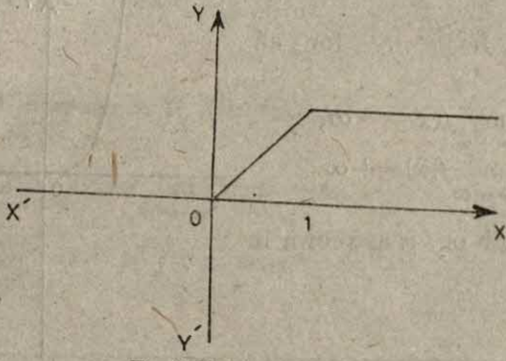


Fig. 2·29 (*a*). Graph of f.

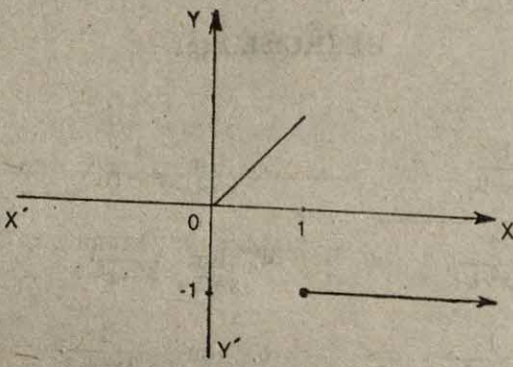


Fig. 2·29 (*b*). Graph of g.

The graph of f can be drawn on paper without lifting the pencil. It has no break. The graph of g has a break at $x=1$. It cannot be drawn without lifting the pencil. In mathematical language we say that f is continuous and g has a discontinuity at $x=1$. Intuitively speaking, continuity of a function means that its graph is all in one piece ; it has no break ; discontinuity (the negation of continuity) means that there is a break in the graph. In the present section we shall try to make this distinction precise.

2·9·1. Continuity at a point

Let a function f be defined on a domain D containing an open interval $I=]a, b[$ and let c be a point of I. f is said to be *continuous* at $x=c$ if $\lim\limits_{x\to c} f(x)$ exists and equals $f(c)$. In other words, *f is continuous at $x=c$ if and only if*

$$\lim_{x\to c+} f(x)=\lim_{x\to c-} f(x)=f(c). \qquad ...(1)$$

If f is not continuous at $x=c$, we say that it is discontinuous at $x=c$.

From (1) we find that to examine a function f for continuity at $x=c$, we have to compute the left-hand limit and right-hand limit of f at $x=c$. If both these limits exist and are equal to the value of f at $x=c$, then f is a continuous at $x=c$, otherwise it is discontinuous at $x=c$. Sometimes it may happen that $\lim_{x\to c} f(x)$ may exist but may not be equal to $f(c)$. In such a case, the function is said to have a removable discontinuity at $x=c$, for the function can be modified by altering the value at $x=c$ so as to get a function which is continuous at $x=c$.

2·9·2. Continuity in an open interval

With the same notation as above, we say that f is continuous on I if it is continuous at every point of I.

2·9·3. Continuity on a closed interval

Let f be defined on a closed interval $[a, b]$. f is said to be continuous at the left-hand end point 'a' if $\lim_{x\to a+} f(x)$ exists and equals $f(a)$. Also f is said to continuous at the right-hand end point b if the left-hand limit at $x=b$, namely $\lim_{x\to b-} f(x)$ exists and equals $f(b)$.

f is said to be continuous on $[a, b]$ if it is continuous at every point of $[a, b]$.

More generally, a real function $f: D\to\mathbf{R}$ is said to be continuous on D if it is continuous at every point of its domain. From our discussion of limits we can get several examples of continuous functions, as well as of functions which fail to be continuous.

Illustrations

1. *A constant function.* Let k be any real number and let f be defined on $\mathbf{R}$ by setting $f(x)=k$ for all $x\in\mathbf{R}$. f is continuous on $\mathbf{R}$. For, if c be any point of $\mathbf{R}$, then $\lim_{x\to c} f(x)=k$, because $f(x)=k$ for all $k\in\mathbf{R}$. Also $f(c)=k$, and consequently $\lim_{x\to c} f(x)=f(c)$.

2. *The identity function.* Let f the function defined on $\mathbf{R}$ by setting $f(x)=x$ for all $x\in\mathbf{R}$. f is continuous on $\mathbf{R}$. For, if c be any point of $\mathbf{R}$, then $\lim_{x\to c} f(x)=\lim_{x\to c} x=c=f(c)$, which implies that f is continuous at $x=c$.

3. Let f be defined on $\mathbf{R}$ by setting $f(x)=x^2$ for all $x\in\mathbf{R}$. f is continuous on $\mathbf{R}$, for if c be any point of $\mathbf{R}$, then

$\lim_{x \to c} f(x) = \lim_{x \to c} x^2 = c^2 = f(c)$, which implies that f is continuous at $x = c$.

2·9·4. Algebra of continuous functions

We already know as to how we can construct new functions from given functions by the four fundamental operations. Continuity is generally preserved under such operations. The following result is found to be very useful in this regard. The proof is a direct consequence of the corresponding results regarding limits, and is therefore omitted.

Theorem 2·9. *Let f and g of be two real functions with a common domain D, let k be any real number, and let f and g be continuous at a point $c \in D$. Then the functions $f+g$, kf, fg, $|f|$ are continuous at a point $x = c$, Furthermore, if $g(c) \neq 0$, then f/g is also continuous at $x = c$.*

2·9·5. Composite of two continuous functions

Theorom 2.10. *Let f and g be two real functions such that range of f is contained in the domain of g. If f is continuous at a point $x = c$, and g is continuous at the point $f(c)$, then the function $g \circ f$ is continuous at c.*

Proof. Since f is continuous at $x = c$, therefore $f(x) \to f(c)$ as $x \to c$. Also, since g is continuous at $f(c)$, therefore $g(t) \to g(f(c))$ as $t \to f(c)$. Now $\lim_{x \to c} (g \circ f)(x) = \lim_{x \to c} g(f(x)) = \lim_{t \to f(c)} g(t) = g(f(c))$ $= (g \circ f)(c)$, where we have written t for $f(x)$ and used the fact that when $x \to c$, $f(x) \to f(c)$. Since $\lim_{x \to c} (g \circ f)(x) = (g \circ f)(c)$, therefore $g \circ f$ is contiouous at $x = c$.

2·9·6. Continuity of polynomials

Let f be a polynomial with real co-efficients and let $c \in \mathbf{R}$. We know that

$$\lim_{x \to c} f(x) = f(c),$$

from which it follows that *every polynomial function is continuous.* The above result is important in as much as it establishes the continuity of a very large class of real functions. It is useful in another way as well. It helps us in establishing the continuity of a function which agrees with a polynomial in some open interval.

Let f be a function defined on a domain D. *Suppose f agrees with a polynomial g on some open interval* $I \subset D$. (This means that $f(x) = g(x)$ for all $x \in I$). *Then f must be continuous on* I. The proof of this fact is rather simple. Suppose $c \in I$. Since g is a polynomial, it is continuous at $x = c$, *i.e.*,

$$\lim_{x \to c} g(x) = g(c) \qquad \ldots (i)$$

Also $g(x)=f(x)$ on I, and therefore for all points close to c, we have

$$g(x)=f(x), \qquad \ldots(ii)$$

and

$$g(c)=f(c). \qquad \ldots(iii)$$

From (i), (ii) and (iii), we find that

$$\lim_{x\to c} f(x)=f(c),$$

i.e., f is continuous at $x=c$. Since c is *any* point of I, it follows that f is continuous on I.

The following examples illustrates how the above result can be applied to discuss continuity of functions.

Example 18. *Let f be defined on $[0, 1]$ by setting*

$$f(0)=0,\ f(x)=\tfrac{1}{2}-x \text{ if } 0<x<\tfrac{1}{2},$$

$$f(\tfrac{1}{2})=\tfrac{1}{2} \text{ and } f(x)=\tfrac{3}{2}-x \text{ if } \tfrac{1}{2}<x\leqslant 1.$$

Show that f is discontinuous at $x=0$ and $x=\frac{1}{2}$, but is continuous at all other points. Draw a rough sketch of the graph of f.

Solution. Since f agrees with the polynomial $\frac{1}{2}-x$ on the open interval $]\,0, \frac{1}{2}[$ therefore f is continuous on $]\,0, \frac{1}{2}[$. Similarly f agrees with the polynomial $\frac{3}{2}-x$ on the open interval $]\,\frac{1}{2}, 1\,[$. Therefore f is continuous on $]\,\frac{1}{2}, 1[$. To examine the continuity at the points $x=0, \frac{1}{2}, 1$, we proceed as follows :

(i) **Continuity at $x=0$.**

$$\lim_{x\to 0+} f(x)=\lim_{x\to 0+} (\tfrac{1}{2}-x)=\tfrac{1}{2},$$

which is not equal to $f(0)$. Therefore f is discontinuous at the left-hand point $x=0$.

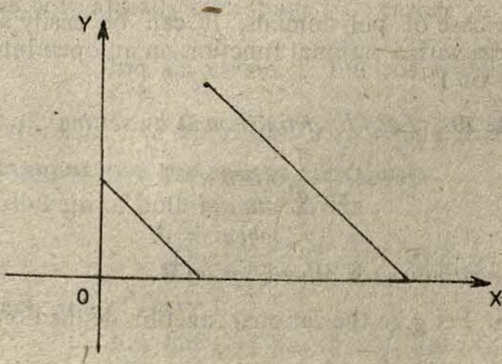


Fig. 2.30. Graph of f.

(*ii*) **Continuity at $x=\frac{1}{2}$.**

$$\lim_{x\to\frac{1}{2}-} f(x)=\lim_{x\to\frac{1}{2}-} (\tfrac{1}{2}-x)=0,$$

$$\lim_{x\to\frac{1}{2}+} f(x)=\lim_{x\to\frac{1}{2}+} (\tfrac{3}{2}-x)=1.$$

Also $$f(\tfrac{1}{2})=\tfrac{1}{2}.$$

We find that $\lim_{x\to\frac{1}{2}+} f(x)$, $\lim_{x\to\frac{1}{2}-} f(x)$, and $f(\frac{1}{2})$ are all different and consequently f is discontinuous at $x=\frac{1}{2}$.

(*iii*) **Continuity at $x=1$.**

$$\lim_{x\to 1-} f(x)=\lim_{x\to 1-} (\tfrac{3}{2}-x)=\tfrac{3}{2}-1=\tfrac{1}{2}.$$

Also $$f(1)=\tfrac{1}{2}.$$

Since $\lim_{x\to 1-} f(x)=f(1)$, therefore f is continuous at $x=1$.

Thus f is discontinuous at $x=0$, $\frac{1}{2}$ and is continuous at every other point. The graph of f is as shown in Fig. 2.30.

2·9·7. Continuity of rational functions

Let $$f(x)=\frac{p(x)}{q(x)},$$

where $p(x)$ and $q(x)$ are polynomials with real coefficients. Let D be the set of points of $\mathbf{R}$ at which $q(x)\neq 0$. As we have seen in (XI) on page 100, for all $c\in D$, we have

$$\lim_{x\to c} f(x)=\frac{p(c)}{q(c)}=f(c),$$

and therefore f is continuous at $x=c$. Thus we find that *a rational function is continuous at every point of its domain.*

As in the case of polynomials, it can be easily seen that if a function f agrees with a rational function on an open interval I, then f is continuous on I.

Example 19. *Let f be defined on* $\mathbf{R}$ *by setting*

$$f(x)=\frac{x^2+1}{x-4}, \text{ when } x\neq 4.$$

$$0 \quad, \text{ when } x=4.$$

Examine f for continuity (or otherwise) on $\mathbf{R}$.

Solution. Let g be the rational function defined on $D=\mathbf{R}\sim\{4\}$ by setting

$$g(x)=\frac{x^2+1}{x-4}.$$

Clearly g is continuous on D. Also f agrees with g on each of the two open intervals $]-\infty, 4[$ and $]4, \infty[$ and is therefore continuous at all points of each of these intervals. It only remains to examine the continuity of f at $x=4$.

$$\lim_{x \to 4-} \frac{x^2+1}{x-4} = \lim_{\substack{h \to 0 \\ h>0}} \frac{(4-h)^2+1}{(4-h)-4};$$

$$= \lim_{\substack{h \to 0 \\ h>0}} \frac{17-8h+h^2}{-h},$$

$$= -\infty,$$

$$\lim_{x \to 4+} \frac{x^2+1}{x-4} = \lim_{\substack{h \to 0 \\ h>0}} \frac{(4+h)^2+1}{(4+h)-4},$$

$$= \lim_{\substack{h \to 0 \\ h>0}} \frac{17+8h+h^2}{h},$$

$$= +\infty.$$

Since $\lim_{x \to 4-} f(x) \neq \lim_{x \to 4+} f(x)$, therefore f is discontinuous at $x=4$.

Thus f is discontinuous at $x=4$ but is continuous at every other point of $\mathbf{R}$.

2·9·8. Continuity of trigonometric functions

We shall now examine the trigonometric functions for continuity. Let us consider each one of them.

A. The sine and cosine functions

As we have already seen (page 107) earlier in this chapter, for all $x \in \mathbf{R}$,

$$\lim_{x \to c} \sin x = \sin c, \quad \lim_{x \to c} \cos x = \cos c.$$

Therefore *the sine and cosine functions are continuous on* $\mathbf{R}$.

B. The tangent and the secant functions

The domain of both tan and sec is

$$D = \left\{ (2n+1)\frac{\pi}{2} : n \in \mathbf{Z} \right\}.$$

At each point $c \in D$,

$$\lim_{x \to c} \tan x = (\lim_{x \to c} \sin x)/(\lim_{x \to c} \cos x),$$

by the quotient theorem on limits

$$=\frac{\sin c}{\cos c}, \text{ by the continuity of sin and cos}$$

$$=\tan c,$$

showing that the tangent function is continuous at $x=c$.

Similarly,

$$\lim_{x\to c} \sec x = \lim_{x\to c} \frac{1}{\cos x} \qquad \text{(why ?)}$$

$$=\frac{1}{\lim\limits_{x\to c} \cos x} \qquad \text{(why ?)}$$

$$=\frac{1}{\cos c}$$

$$=\sec c,$$

show that the secant function is continuous at $x=c$.

Thus *the cosine and the secant functions are continuous at all points of their domain.*

C. The cotangent and the cosecant functions

The domain of both cot and csc is $D^*=\{n\pi : n\in \mathbf{Z}\}$.

At each point $c\in D^*$,

$$\lim_{x\to c} \cot x = (\lim_{x\to c} \cos x)/(\lim_{k\to c} \sin x) \qquad \text{(why ?)}$$

$$=\frac{\cos c}{\sin c}, \qquad \text{(why ?)}$$

$$=\cot c.$$

$$\lim_{x\to c} \csc x = 1/\lim_{x\to c} \sin x,$$

$$=1/\sin c,$$

$$=\cos c.$$

From the above we find $\cot x$ and $\csc x$ are both continuous at $x=c$.

Thus *the cotagent and the cosecant functions are continuous at all points of their domain.*

Remark. In (B) and (C) above we could as well have used the quotient theorem for continuous functions (which of course depends on the quotient theorem for limits).

Example 20. *Examine for continuity, the function f defined by*

$$f(x)=\begin{cases} \dfrac{\sin x}{x}, & \text{when } x\neq 0. \\ 0, & \text{when } x=0. \end{cases}$$

Solution.

Case I. When $x \neq 0$.

$$f(x)=\frac{p(x)}{q(x)},$$

where $p(x)=\sin x$, $q(x)=x$. The functions p and q are both continuous on $\mathbf{R}$ and q does not vanish whenever $x \neq 0$. Therefore by the quotient theorem for continuous functions, f is continuous whenever $x \neq 0$.

Case II. $x=0$.

$$\lim_{x \to 0} \frac{\sin x}{x}=1 \neq f(0).$$

Therefore f is discontinuous at $x=0$. Thus we find that the given function is continuous on $\mathbf{R}$ except at $x=0$ where it is discontinuous.

Remark. The graph of f is as shown in Fig. 2·31. Observe that it crosses the x-axis at $\pm\pi$, $\pm 2\pi$, ...

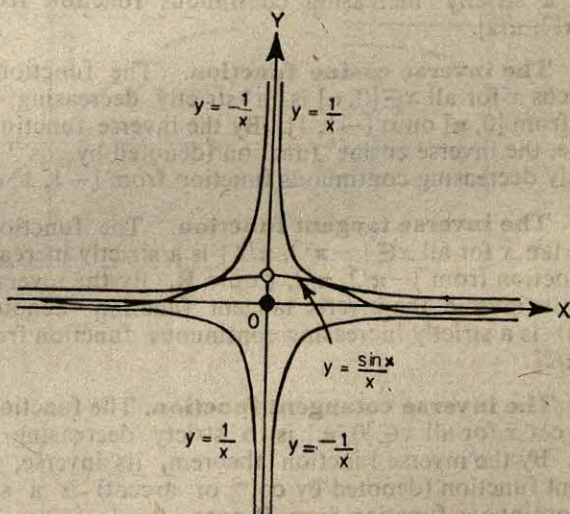


Fig. 2·31. Graph of $\frac{\sin x}{x}$

Example 21. *Show that the function $f(x)=|\sin x|$ is continuous everywhere.*

Solution. Let u be the function $x \to \sin x$, and let v be the function $x \to |x|$. Both u and v are continuous on $\mathbf{R}$. Since the

composite of two continuous functions is continuous, therefore the function vou is continuous. But

$$(vou)(x)=v(u(x))=v(\sin x)=|\sin x|f(x)$$

for all $x\in\mathbf{R}$. Therefore f is continuous on $\mathbf{R}$.

2·9·9. Continuity of inverse trigonometric functions

Having considered the continuity of trigonometric functions we shall now examine the inverse trigonometric functions for continuity. In order to do so, we need the following :

Theorem 2·11. (*Inverse function theorem for continuous functions*). *If $f : D\to\mathbf{R}$ is a strictly monotone continuous function with domain D and range $\mathbf{R}$, then f possesses an inverse $f^{-1} : \mathbf{R}\to D$ which is a strictly monotone continuous function with domain $\mathbf{R}$ and range D.*

We omit the proof of the above theorem.

(*a*) **The inverse sine function.** The function f defined by $f(x)=\sin x$ for all $x\in[-\pi/2, \pi/2]$ is a strictly increasing continuous function from $[-\pi/2, \pi/2]$ onto $[-1, 1]$. By the inverse function theorem its inverse, the inverse sine function (denoted by $\sin^{-1}$ or arcsin) is a strictly increasing continuous function from $[-1, 1]$ onto $[-\pi/2, \pi/2]$.

(*b*) **The inverse cosine function.** The function g defined by $f(x)=\cos x$ for all $x\in[0, \pi]$ is a strictly decreasing continuous function from $[0, \pi]$ onto $[-1, 1]$. By the inverse function theorem, its inverse, the inverse cosine function (denoted by $\cos^{-1}$ or arccos) is a strictly decreasing continuous function from $[-1, 1]$ onto $[0, \pi]$.

(*c*) **The inverse tangent function.** The function h defined by $h(x)=\tan x$ for all $x\in]-\pi/2, \pi/2[$ is a strictly increasing continuous function from $]-\pi/2, \pi/2,[$ onto $\mathbf{R}$. By the inverse function theorem, its inverse, the inverse tangent function (denoted by $\tan^{-1}$ or arctan) is a strictly increasing continuous function from R onto $]-\pi/2, \pi/2[$.

(*d*) **The inverse cotangent function.** The function u defined by $u(x)=\cot x$ for all $x\in]0, \pi[$ is a stricty decreasing continuous function. By the inverse function theorem, its inverse, the inverse contangent function (denoted by $\cot^{-1}$ or arccot) is a strictly decreasing continuous function from $\mathbf{R}$ onto $]0, \pi]$.

(*e*) **The inverse secant function.** The function v defined by $v(x)=\sec x$ for all $x\in[0, \pi/2\ [\cup]\ \pi/2, \pi]$ is a continuous function which is strictly increasing on each of the intervals $]-\infty, -1]$ and $]1, \infty[$ onto $\{x : x\in\mathbf{R}, |x|\geqslant 1\}$. By the inverse function theorem, its inverse, the inverse secant function (denoted by $\sec^{-1}$ or arcsec) is a continuous function from $\{x : x\in\mathbf{R}, |x|\geqslant 1\}$ onto $[0, \pi/2[\cup]\pi/2, \pi[$, which is strictly increasing in each of the intervals $]-\infty, -1]$ and $]1, \infty[$.

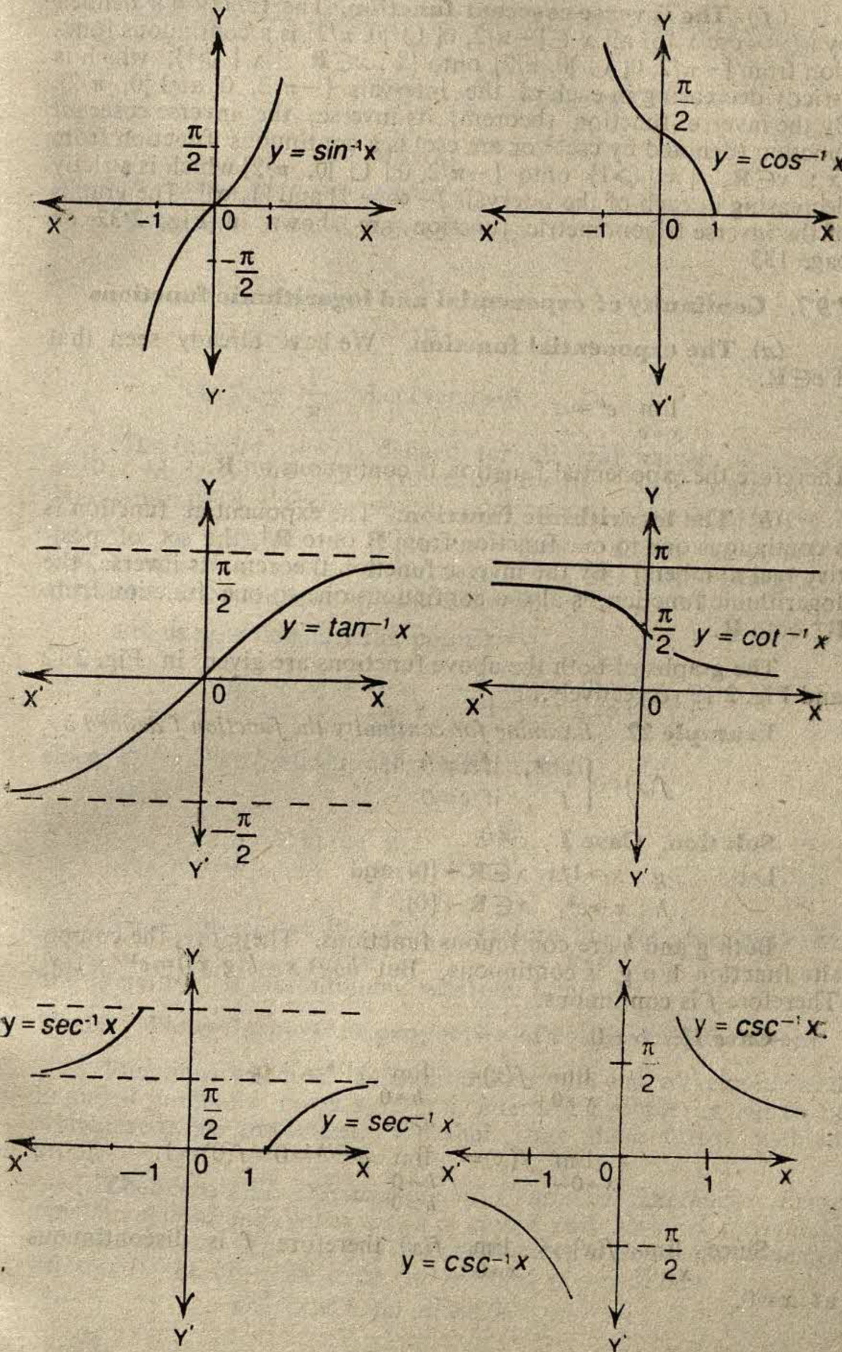

y = sin⁻¹x
y = cos⁻¹x
y = tan⁻¹x
y = cot⁻¹x
y = sec⁻¹x
y = sec⁻¹x
y = csc⁻¹x
y = csc⁻¹x
π/2
−π/2
π
X
X'
Y
Y'
0
1
−1

Fig. 2·32.

(*f*) **The inverse cosecant function.** The function w defined by $w(x)=\csc x$ for all $x \in [-\pi/2, 0[\cup]0, \pi/2]$ is a continuous function from $[-\pi/2, 0[\cup]0, \pi/2]$ onto $\{x : x\in \mathbf{R}, \mid x \mid \geqslant 1\}$, which is strictly decreasing in each of the intervals $[-\pi/2, 0[$ and $]0, \pi/2]$. By the inverse function theorem, its inverse, the inverse cosecant function (denoted by $\csc^{-1}$ or arc csc) is a continuous function from $\{x : x\in\mathbf{R}, \mid x \mid \geqslant 1\}$ onto $[-\pi/2, 0 [\cup]0, \pi/2]$ which is strictly decreasing in each of the intervals $]-\infty, -1]$ and $[1, \infty[$. The graphs of the inverse trigonometric function are shown in Fig. 2·32 on page 133.

2·9·7. Continuity of exponential and logarithmic functions

(*a*) **The exponential function.** We have already seen that if $c\in\mathbf{R}$.

$$\lim_{x\to c} e^x = e^c.$$

Therefore the exponential function is continuous on **R**.

(*b*) **The logarithmic function.** The exponential function is a continuous one-to-one function from **R** onto $\mathbf{R}^+$ (the set of positive real numbers) By the inverse function theorem, its inverse, the logarithmic function, is also a continuous one-to-one function from $\mathbf{R}^+$ onto **R**.

The graphs of both the above functions are given in Fig. 2·12 and Fig. 2·13 respectively.

Example 22. *Examine for continuity the function f defined by*

$$f(x)=\begin{cases} e^{1/x}, & \text{if } x\neq 0 \\ 1, & \text{if } x=0. \end{cases}$$

Solution. Case I. $x\neq 0$.

Let $g : x\to 1/x,\ x\in\mathbf{R}\sim\{0\}$, and

$h : x\to e^x,\ x\in\mathbf{R}\sim\{0\}$.

Both g and h are continuous functions. Therefore, the composite function h o g is continuous. But $(hog)\, x=h(g(x))=e^{1/x}=f(x)$. Therefore f is continuous.

Case II. $x=0$.

$$\lim_{x\to 0+} f(x)= \lim_{\substack{h\to 0\\ h>0}} e^{1/h}=+\infty$$

$$\lim_{x\to 0-} f(x)= \lim_{\substack{h\to 0\\ h>0}} e^{-1/h}=0=f(0)=1.$$

Since $\lim\limits_{x\to 0+} f(x) \neq \lim\limits_{x\to 0^-} f(x)$, therefore f is discontinuous at $x=0$.

Example 23. *Examine for continuity the function f defined on* **R** *by*

$$f(x)=\frac{e^{1/x}}{1+e^{1/x}},\ \text{if } x\neq 0$$

$$f(0)=k,$$

where k is some fixed real number.

Solution. Consider the functions

$$u : x\to \frac{x}{1+x},\ x>0$$

$$v : x\to e^{x},\ \forall\ x\in \mathbf{R}$$

$$w : x\to \frac{1}{x},\ \text{whenever } x\neq 0.$$

The function $v\circ w$ is defined for all real values of x except $x=0$. Also, $(v\circ w)(x)>0$ for all $x\in\mathbf{R}\sim\{0\}$. Since v and w are both continuous, therefore $v\circ w$ is continuous. Also, $(v\circ w)(x)=e^{1/x}$.

Let us consider the function $u\circ(v\circ w)$. Clearly $[u\circ(v\circ w)](x)=f(x)$, whenever $x\neq 0$. Since u and $v\circ w$ are both continuous, therefore f is continuous for all $x\in\mathbf{R}\sim\{0\}$.

Let us now consider the point $x=0$.

$$\lim_{x\to 0+} f(x)=\lim_{\substack{h\to 0\\ h>0}}\frac{e^{1/h}}{1+e^{1/h}}=\lim_{\substack{h\to 0\\ h>0}}\frac{1}{e^{-1/h}+1}=1,$$

since $e^{-1/h}\to 0$ as $h\to 0$ through positive values.

Also

$$\lim_{x\to 0-} f(x)=\lim_{\substack{h\to 0\\ h>0}}\frac{e^{-1/h}}{1+e^{-1/h}}=0.$$

Since $\lim\limits_{x\to 0^+} f(x)\neq \lim\limits_{h\to 0-} f(x)$, therefore $\lim\limits_{h\to 0} f(x)$ does not exist, and therefore f is discontinuous whatever k may be.

2·9·11. Three important properties of continuous functions

Functions which are defined and continuous on closed and bounded intervals possess several interesting properties, three of which are rather important. We shall state them below without proof.

Theorem 2·12. (*Boundness of Continuous functions*). *Every function defined and continuous on a closed and bounded interval is bounded therein. That is, if f is continuous on a closed and bounded interval I(=[a, b]), there exist real numbers u, v such that*

$$u\leqslant f(x)\leqslant v,\ \textit{for all } x\in I.$$

The above theorem says, in effect, that the range of a continuous function defined on a closed and bounded interval is contained in a closed are bounded interval. If any of the hypotheses of the above theorem are violated, the conclusion may fail to hold. For example, consider the function f defined by setting :

(*a*) $f(x)=x$ for all x on $[0, \infty[$

(*b*) $f(0)=0, f(x)=\frac{1}{x}$ for $x \in]0, 1[$

(*c*) $f(x)=\frac{1}{x}$ on $]0, 1]$

In (*a*), the domain of f is the unbounded interval $[1, \infty[$. [see Fig. 2·33(*a*)].

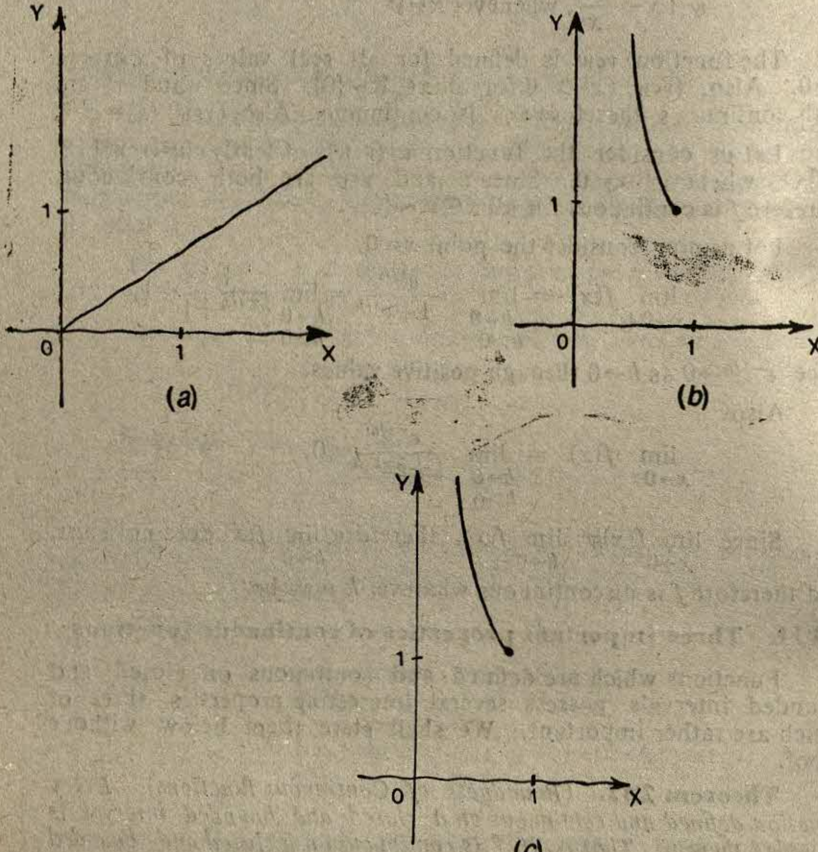


Fig. 2,33.

Observe that here f is continuous but the domain of f is not bounded.

In (b), f is discontinuous at $x=0$ [see fig. 2.33 (b)].

In (c), the domain of f is not closed.

In all the three cases f is unbounded so that the conclusion of theorem 2.12 fails to hold.

The graph of f is as shown in Fig. 2.34.

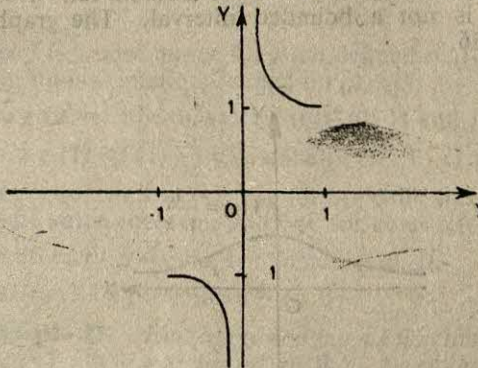


Fig. 2·34.

Reason. f is not continuous at $x=0$, for $\lim_{x\to 0+} f(x)=+\infty$

$\lim_{x\to 0-} f(x)=-\infty, f(0)=1.$

Theorem 2·13. (*Attainment of supremum and infimum*). *Every function defined and continuous on a closed and bounded interval attains its supremum and infimum. That is, if f is defined and continuous on a closed and bounded interval $I(=[a, d])$, if u be the supremum of f on I and l be the infimum of f on I, then there exist p and q in I, such that $f(u)=p, f(l)=q$.*

If any of the hypotheses of theorem 2.13 are violated, the conclusion of the theorem may fail to hold. This can be seen by considering the following examples :

(*a*) Let f be the function defined by setting

$f(x)=x$ for all x in $[0, 1[$.

f is continuous in $[0, 1[$, it is bounded above in $[0, 1[$, sup. f being 1, but there is no point of the domain of f at which this supremum is attained, that is there is no point of $[0, 1[$, at which $f(x)=1$.

Here, the domain of f is not a closed interval, so that one of the hypotheses of the theorem is violated.

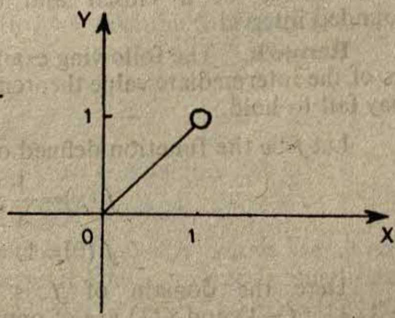


Fig. 2.35.

(b) Let f be the function defined by $f(x)=\frac{1}{1+x^2}$, for all $x \in \mathbf{R}$. f is clearly continuous on $\mathbf{R}$, it is bounded below as well as above, sup. f being 1 and inf. f being 0. While sup. f is attained at $x=0$ $(f(0)=1=\text{sup. } f)$, inf. f is not attained, that is, there is no point of $\mathbf{R}$ at which $f(x)=\text{inf. } f$ (Where does the theorem fail ?). Observe that the domain of f is not a bounded interval. The graph of f is as shown in Fig. 2.36.

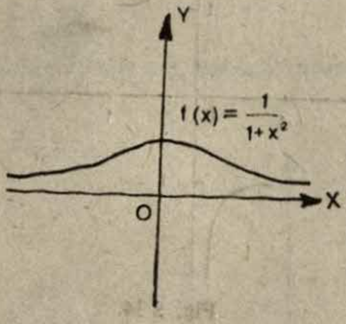


Fig. 2·36.

Theorem 2·14. (*Intermediate Value Theorem.*) *If f be continuous on the closed and bounded interval* $[a, b]$, *and c be any real number between $f(a)$ and $f(b)$, then there exists a real number x_0 in* $]a, b[$ *such that $f(x_0)=c$.*

The above theorem says in effect that if a function is continuous on a closed and bounded interval $[a, b]$, then it takes every value lying between its bounds. Equivalently, it says that if f is continuous on $[a, b]$, then it assumes every value lying between any two of its values. Expresed differently, it says that the continuous image of an interval is an interval.

Theorems 2·13 and 2·14, when put together assert that a continuous image of a closed and bounded interval is a closed and bounded interval.

Remark. The following example shows that if the hypotheses of the intermediate value theorem are violated, then its conclusion may fail to hold.

Let f be the function defined on $[-1, 1]$ by setting

$$f(x)=\frac{1}{x}, \text{ when } x \neq 0,$$

$$f(0)=1.$$

Here the domain of f is the closed and bounded interval $[-1, 1]$, $f(-1)$ and $f(1)$ are of opposite signs, but $f(x)$ does not vanish.

Example 24. *Let f be continuous on $[0, 1]$ and let $f(x)$ lie in $[0, 1]$ for each x in $[0, 1]$. Prove that $f(x)=x$ for some x in $[0, 1]$.*

Solution. If $f(0)=0$, or $f(1)=1$, then the result is obvious. Let us therefore consider the case when $f(0)\neq 0$ and $f(1)\neq 1$.

Consider the function

$$g(x)=f(x)-x.$$

Since f is a continuous function defined on $[0, 1]$, therefore g is also a continuous function defined on $[0, 1]$.

Also $g(0)=f(0)>0$, since $f(0)\in[0, 1]$ and $f(0)\neq 0$.

Again $g(1)=f(1)-1<0$, since $f(1)\in[0, 1]$ and $f(1)\neq 1$.

Since $g(0)$ and $g(1)$ are of opposite signs, therefore by the intermediate value theorem, $g(x)=0$ for some $x\in]0, 1[$,

Therefore $f(x)-x=0$, for some $x\in]0, 1[$.

Thus $f(x)=x$ for some $x\in]0, 1[$.

Example 25. *How many continuous functions are there on* **R** *which satisfy $[f(x)]^2=x^2$ for all $x\in\mathbf{R}$? Draw the graphs of all such functions.*

Solution. Since $[f(x)]^2=x^2$ for all $x\in\mathbf{R}$, therefore

$$f(0)=0. \qquad ...(i)$$

Let us now consider the form of f on the interval $I=]0, \infty[$. Since f is continuous on I, and $f(x)\neq 0$ (because) $[f(x)]^2\neq 0$ for any $x\in I$, therefore by the intermediate value theorem either

$$f(x)=x \text{ for all } x\in I.$$

or

$$f(x)=-x \text{ for all } x\in I. \qquad ...(ii)$$

Next, let us consider the form of f on the interval $J=]-\infty, 0[$.

Since f is continuous on J, and $f(x)\neq 0$ (because) $[f(x)^2]\neq 0$ for any $x\in J$, therefore by the intermediate value theorem either $f(x)=x$ for all $x\in J$ or $f(x)=-x$ for all $x\in J$. ...(iii)

Combining (i), (ii) and (iii) we have the following four possibilities :

(i) $f(x)=x$ for all $x\in\mathbf{R}$.

(ii) $f(x)=-x$ for all $x\in\mathbf{R}$.

(iii) $f(x)=x$ when $x\geqslant 0$, $f(x)=-x$ when $x<0$.

(iv) $f(x)=-x$ when $x\geqslant 0$, $f(x)=x$ when $x<0$.

Since the functions described in (i)—(iv) above are all continuous, therefore we find that there are four continuous functions on **R** satisfying $[f(x)]^2=x^2$ for all $x\in\mathbf{R}$, namely $f(x)=x$, $f(x)=-x$ $f(x)=|x|$, $f(x)=-|x|$.

The graphs of these functions are as shown in Fig. 2.37.

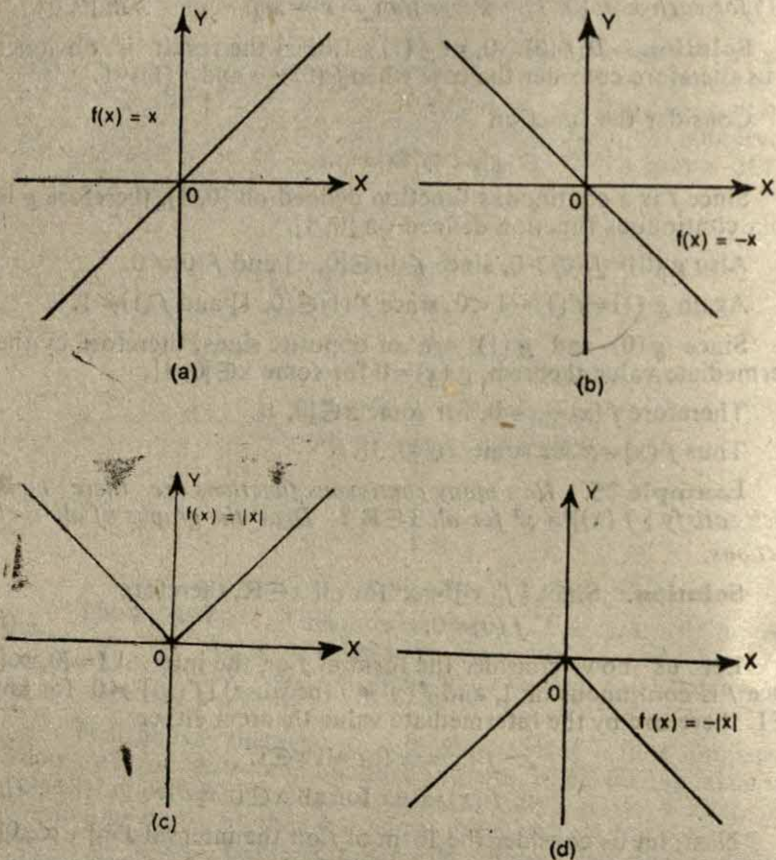


Fig. 2.37.

Example 26. *Let f be a continuous function on* $[-1, 1]$ *such that* $[f(x)]^2+x^2=1$, *for all* $x\in[-1, 1]$.

Show that either $f(x)=\sqrt{1-x^2}$ *for all x in* $[-1, 1]$ *or* $f(x)=-\sqrt{1-x^2}$ *for all x in* $[-1, 1]$.

Solution. Since $[f(x)]^2+x^2=1$, therefore $f(x)=\pm\sqrt{1-x^2}$, *i.e.*, if $x\in[-1, 1]$, then either $f(x)=\sqrt{1-x^2}$ or $f(x)=-\sqrt{1-x^2}$. What we are required to show is that

either $f(x)=\sqrt{1-x^2}$ for all $x\in[-1, 1]$,

or $\quad f(x)=-\sqrt{1-x^2}$ for all $x\in[-1, 1]$.

Suppose, if possible, that there exist two points, x_1, x_2 in $]-1, 1[$, such that

$$f(x_1)=-\sqrt{1-x_1^2}, f(x_2)=\sqrt{1-x_2^2}.$$

Without loss of generality, let $x_1<x_2$.

Now f is continuous in $[x_1\ x_2]$, and $f(x_1), f(x_2)$ are of opposite signs. By the intermediate value theorem, there must exist

$$c \in]x_1, x_2[\subset]-1, 1[, \text{ such that } f(c)=0.$$

But this is impossible, since $f(c)=0$, iff $c^2=1$ (because $(f(c))^2+c^2=1$).

Hence either $f(x)=\sqrt{1-x^2}$, for all $x\in[-1, 1]$,

or $\qquad f(x)=-\sqrt{1-x^2}$, for all $x\in[-1, 1]$.

Example 27. *A function f is continuous in the interval $[0, 1]$ and assumes only rational values in the entire interval. If $f(x)=\frac{1}{2}$ when $x=\frac{1}{2}$, prove that $f(x)=\frac{1}{2}$ everywhere.*

Solution. Suppose, if possible, that there exists a point $c\in[0, 1]$ such that $f(c)\neq\frac{1}{2}$. It is obvious that $0\not<\frac{1}{2}$. By the law of trichotomy, either $c<\frac{1}{2}$ or $\frac{1}{2}<c$. Without loss of generality, let us assume that $c<\frac{1}{2}$. The function f is continuous in $[c, \frac{1}{2}]$ and therefore by the intermediate value theorem it must take every value lying between $f(c)$ and $f(\frac{1}{2})$. But this is not possible, because $f(c)$ and $f(\frac{1}{2})$ are two distinct rational numbers between which there lie infinitely many irrational numbers, and $f(x)$ does not take any irrational value. The contradiction shows that there does not exist any $c\in[0, 1]$ such that $f(c)$ is different from $\frac{1}{2}$.

Hence $f(x)=\frac{1}{2}$ everywhere.

EXERCISE 2 (*h*)

1. Show that each of following functions is continuous for all $x\in\mathbf{R}$:
   (*a*) $f(x)=x^2+1$.
   (*b*) $f(x)=x^3-2$.
   (*c*) $f(x)=x^4-7x+5$.
2. Show that the function f defined on $\mathbf{R}$ by setting $f(x)=1/x$ if $x\neq0, f(0)=0$ is discontinuous at $x=0$ but is continuous at every other point.
3. Examine the function f, defined on $\mathbf{R}$ by setting
   $$f(x)=\begin{cases}-x^2, \text{ if } x\leqslant0\\ 5x-4, \text{ if } 0<x<1\\ 4x^2-3, \text{ if } x\geqslant1\end{cases}$$
   for points of discontinuity.
4. Let f be defined by setting
   $$f(x)=\frac{\sin 2x}{5x}, \text{ if } x\neq0$$
   $$f(0)=k.$$

For what value of k is f continuous at $x=0$?

5. Examine for continuity at $x=0$, the function f defined on $\mathbf{R}$ in each of the following cases :

(i) $f(x)=\begin{cases} \dfrac{\tan 2x}{x}, & \text{when } x\neq 0 \\ 2, & \text{when } x=0. \end{cases}$

(ii) $f(x)=\begin{cases} \dfrac{x}{\sin 3x}, & \text{when } x\neq 0 \\ 1, & \text{when } x=0. \end{cases}$

(iii) $f(x)=\begin{cases} \sin \dfrac{1}{x}, & \text{when } x\neq 0. \\ 0, & \text{when } x=0. \end{cases}$

(iv) $f(x)=\begin{cases} \cos \dfrac{1}{x}, & \text{when } x\neq 0. \\ 1, & \text{when } x=0. \end{cases}$

(v) $f(x)=\begin{cases} x\sin \dfrac{1}{x}, & \text{when } x\neq 0 \\ 0, & \text{when } x=0. \end{cases}$

6. Examine for continuity at $x=0$, the function f defined on $\mathbf{R}$ in each of the following of cases :

(i) $f(x)=\dfrac{1}{1-e^{1/x}}$, if $x\neq 0$; $f(0)=0$.

(ii) $f(x)=x\dfrac{e^{1/x}-e^{-1/x}}{e^{1/x}+e^{-1/x}}$, if $x\neq 0$, $f(0)=0$.

(iii) $f(x)=\begin{cases} \dfrac{e^{1/x^2}}{1-e^{1/x^2}}, & \text{if } x\neq 0 \\ 1, & \text{if } x=0. \end{cases}$

7. Let f be the function defined on $\mathbf{R}$ by setting $f(x)=[x]$, for all $x\in\mathbf{R}$, where $[x]$ denotes the greatest integer not exceeding x. Show that f is discontinuous at the points $x=0, \pm 1, \pm 2, \pm 3,$ .... and is continuous at every other point.

8. Let f be defined on an interval I and let f be continuous at a point $c\in$ I. Is it necessary that f be continuous at c? Justify your answer.

9. Let f be the function defined on $\mathbf{R}$ by setting $f(x)=x-[x]$, for all $x\in\mathbf{R}$.

Examine f for continuity at the points $x=0, \pm 1, \pm 2, \pm 3, \cdots$

10. Let f be defined on $\mathbf{R}$ by setting

$$f(t)=\begin{cases} t & , \text{ if } 0 \leqslant t < \frac{1}{2}, \\ 0 & , \text{ if } t=\frac{1}{2}, \\ t-1, & \text{if } \frac{1}{2} < t \leqslant 1, \end{cases}$$

and $f(n+t)=f(t)$, where n is any integer. Determine the points of discontinuity of f.

TEST YOUR UNDERSTANDING II

In each of the following problems, four alternatives are given out of which exactly one is correct. Put a tick mark (✓) against the correct alternative.

1. The domain of the function $\dfrac{1}{x^2-1}$ is
   (a) $\{-1, 1\}$ (b) $[-1, 1]$ (c) $]-1, 1[$ (d) $\mathbf{R}\sim\{-1, 1\}$.
2. The range of the function $3 \cos 2x$ is
   (a) $[-1, 1]$ (b) $[-3, 3]$ (c) $[-2, 2]$ (d) $\mathbf{R}$.
3. $\lim\limits_{x\to\pi} \dfrac{\tan x}{\pi-x}$ equals
   (a) 1 (b) -1 (c) $+\infty$ (d) 0.
4. Range of the function $2 ln (x-1)$ is
   (a) 0, (b) $[1, \infty[$ (c) $\mathbf{R}$ (d) $[2, \infty[$.
5. $\lim\limits_{x\to 1+} \dfrac{1}{1-x}$ is
   (a) $-\infty$ (b) $+\infty$ (c) 0 (d) 1.
6. $\lim\limits_{x\to-\infty} e^{1/x}$ is
   (a) $-\infty$ (b) 0 (c) -1 (d) 1.
7. The function f defined by $f(x)=\dfrac{\sin 3x}{\sin 5x}$, when $x\neq 0$.
   $f(0)=k$ is continuous at $x=0$. The value of k is
   (a) 3 (b) $\frac{5}{3}$ (c) $\frac{3}{5}$ (d) 5.
8. The function f is defined by $f(x)=2x+3$ if $x\leqslant 0$,
   $f(x)=ax^2+bx+c$ if $x>0$. The value of c is
   (a) 2 (b) -3 (c) 0 (d) 3.
9. $\lim\limits_{x\to+\infty} -3 ln (x-4)$ is
   (a) $+\infty$ (b) $-\infty$ (c) 0 (d) 1.

10. Let f be the function defined by setting

$$f(x)=\begin{cases} x^2-1 & \text{if } x<2 \\ a-2x^2 & \text{if } x>2 \end{cases}$$

If f is continuous on **R**, the value of a is

(*a*) 1 (*b*) $\frac{1}{2}$ (*c*) -8 (*d*) 11.

REVIEW EXERCISE II

1. Find the domain of the function f defined by $f(x)=\dfrac{1}{x^2+1}$.

2. Find the domain of each of the following functions :

(*a*) $f(x)=\sqrt{x^2-1}$ (*b*) $f(x)=\sqrt{2-x^2}$ (*c*) $f(x)=\dfrac{\sqrt{x^2-1}}{\sqrt{x-2}}$.

3. Find the domain of the function f defined by setting

$f(x)=\sqrt{\dfrac{x}{x+1}}$.

4. Find the range of each of the following functions :

(*a*) $f(x)=\sqrt{x^2-9}$ (*b*) $f(x)=\sqrt{16-x^2}$

(*c*) $f(x)=\sqrt{\dfrac{x+1}{x}}$.

5. Find the domain and range of each of the following functions :

(*a*) $2 \sin 3x$ (*b*) $3 \cos 2x$ (*c*) $4 \cot^2 x$

(*d*) $\frac{1}{2} \csc^{-1} 2x$ (*e*) $6 \tan^{-1} 5x$ (*f*) $-5 \sec^{-1} 4x$.

6. Find the domain and range of each of the following functions :

(*a*) e^{2x} (*b*) $3e^{-4x}$ (*c*) $5e^{|x|}$ (*d*) $ln\,(1/x)$

(*e*) $2 ln\,(1-x)$ (*f*) $-4\, ln\,(1+2x)$.

7. What can you say about the domain of $f: x \to ln\, ln \sin x$? Does it define a function ?

8. Evaluate $\lim\limits_{x\to 2} \dfrac{\sqrt{2x}-2}{x-2}$. (*A.I.S.S.C.E. 1988*)

9. Evaluate $\dfrac{\tan 3x-2x}{3x-\sin^2 x}$. (*A.I.S.S.C.E. 1989*)

10. Evaluate $\lim\limits_{x\to \pi/2} \tan x\, ln\,(\sin x)$.

(*Roorkee Entrance, 1989*)

11. A function f is defined as follows :

$$f(x)=\begin{cases} x^4, & x^2<1 \\ x, & x^2\geqslant 1 \end{cases}$$

Discuss the existence of the limit at $x=1$ and $x=-1$.

12. Find the interval in which the values of $f(x)$ lie where

$f(x)=3\sin\sqrt{\left(\frac{\pi^2}{16}-x^2\right)}$. (*I.I.T. J.E.E., 1983*)

13. Evaluate $\lim_{x\to 1}\frac{x-1}{2x^2+x-3}$. (*I.I.T. J.E.E., 1976*)

14. Evaluate $\lim_{x\to a}\frac{\sqrt{a+2x}-\sqrt{3x}}{\sqrt{3a+x}-2\sqrt{x}}$ $(a\neq 0)$. (*I.I.T. J.E.E., 1978*)

15. Evaluate $\lim_{x\to 0}\frac{\tan x-\sin x}{x^3}$. (*I.I.T. J.E.E., 1974*)

16. Evaluate $\lim_{x\to 0}\frac{\tan 2x-x}{3x-\sin x}$. (*I.I.T. J.E.E., 1971*)

17. Find $\lim_{x\to 1+}\frac{\ln x}{x-1}$. (*I.I.T. J.E.E., 1973*)

18. Find $\lim_{x\to \pi/4}\frac{\sin x-\cos x}{x-\pi/4}$. (*I.I.T. J.E.E., 1971*)

19. Find $\lim_{x\to 1}(1-x)\tan\frac{\pi x}{2}$. (*I.I.T. J.E.E., 1978*)

20. Let $f(x)$ be a continuous function and let $g(x)$ be a discontinuous function. Prove that $f(x)+g(x)$ is a discontinuous function. (*I.I.T. J.E.E., 1987*)

21. Let $f(x)=\begin{cases}1+x, & 0\leqslant x\leqslant 2\\ 3-x, & 2<x\leqslant 3\end{cases}$

Determine the form of $g(x)=f(f(x))$ and hence find the points of discontinuity of g, if any. (*I.I.T. J.E.E., 1983*)

22. Find the domain and range of $f(x)=\frac{x^2}{1+x^2}$. Is the function one-to-one? (*I.I.T. J.E.E., 1978*)

23. Find the domain of the function

$f(x)=\frac{1}{\log_{10}(1-x)}-\sqrt{x+2}$. (*I.I.T. J.E.E., 1983*)

24. Use the formula $\lim_{x\to 0}\frac{a^x-1}{x}=\ln a$ to find $\lim_{x\to 0}\frac{2^x-1}{(1+x)^{1/2}-1}$. (*I.I.T. J.E.E., 1982*)

25. Determine the values of a, b, c for which the function

$$f(x)=\begin{cases} \dfrac{\sin (a+1)x+\sin x}{x} & \text{for } x<0 \\ c & \text{for } x=0 \\ \dfrac{(x+b\,x^2)^{1/2}-x^{1/2}}{bx^{3/2}} & \text{for } x>0 \end{cases}$$

is continuous at $x=0$. (*I.I.T. J.E.E., 1982*)

26. The function

$$f(x)=\frac{\ln (1+ax)-\ln (1-bx)}{x}$$

is not defined at $x=0$. Find the value of $f(0)$ so that f is continuous at $x=0$. (*I.I.T. J.E.E., 1983*)

27. Construct the graph of the function given below :

$$f(x)=\begin{cases} x-1, & x<0 \\ \frac{1}{4}, & x=0 \\ x^2, & x>0 \end{cases}$$

Find $\lim\limits_{x\to 0^+} f(x)$ and $\lim\limits_{x\to 0-} f(x)$. Discuss the continuity of $f(x)$ at $x=0$. (*Roorkee Entrance, 1988*)

SUMMARY

1. *Domain* of a function $f : S \to T$ is S, and *range of*

$$f=[\,f(x) : x \in S].$$

2. *Sum, difference, product and quotient of two functions.*

Let f and g be two real functions with the same domain D. Then

$$(f+g)(x)=f(x)+g(x)$$
$$(f-g)(x)=f(x)-g(x)$$
$$(fg)(x)=f(x)g(x)$$
$$(f/g)(x)=f(x)/g(x), \text{ provided } g(x) \text{ is not zero anywhere on } D.$$

3. *Composite of functions*

If $f : X\to Y$, $g : Y\to Z$, then $(gof)(x)=g(\,f(x))$.

4. *Inverse functions*

If $f: X\to Y$, $g : Y\to X$ be both one-to-one onto, such that gof and fog are identity functions on their respective domains, then f^{-1} and g^{-1} both exist, and $f^{-1}=g$, $g^{-1}=f$.

5. $e^x=1+x+\dfrac{x^2}{2!}+\ldots\ldots$ for all $x\in R$.

6. $\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\ldots\ldots$, if $|x|<1$.

7. *Table of domain and range of some standard functions*

Function	*Domain*	*Range*
$\sin x$	$\mathbf{R}$	$[-1, 1]$
$\cos x$	$\mathbf{R}$	$[-1, 1]$
$\tan x$	$\mathbf{R}\sim\{(2n+1)\pi/2 : n\in\mathbf{Z}\}$	$\mathbf{R}$
$\cot x$	$\mathbf{R}\sim\{n\pi : n\in\mathbf{Z}\}$	$\mathbf{R}$
$\sec x$	$\mathbf{R}\sim\{(2n+1)\pi/2 : n\in\mathbf{Z}\}$	$\{x : \|x\|\geqslant 1\}$
$\csc x$	$\mathbf{R}\sim\{n\pi : n\in\mathbf{Z}\}$	$\{x : \|x\|\geqslant 1\}$
$\sin^{-1} x$	$[-1, 1]$	$[-\pi/2, \pi/2]$
$\cos^{-1} x$	$[-1, 1]$	$[0, \pi]$
$\tan^{-1} x$	$\mathbf{R}$	$]-\pi/2, \pi/2[$
$\cot^{-1} x$	$\mathbf{R}$	$]0, \pi[$
$\sec^{-1} x$	$\{x : \|x\|\geqslant 1\}$	$[0, \pi/2[\cup]\pi/2, \pi]$
$\csc^{-1} x$	$\{x : \|x\|\geqslant 1\}$	$[-\pi/2, 0[\cup]0, \pi/2]$
e^x	$\mathbf{R}$	$\mathbf{R}^+$
$\ln x$	$\mathbf{R}^+$	$\mathbf{R}$

8. Let f and g be defined on some open interval $\mathbf{I}$ containing c, but not necessarily at $x=c$. If $\lim\limits_{x\to c} f(x)=l$, $\lim\limits_{x\to c} g(x)=m$, and k be some fixed real number, then

(*a*) $\lim\limits_{x\to c}\,[f(x)+g(x)]=l+m$

(*b*) $\lim\limits_{x\to c}\,[kf(x)]=kl$

(*c*) $\lim\limits_{x\to c}\,[f(x)g(x)]=lm$

(*d*) $\lim\limits_{x\to c}\,[f(x)/g(x)]=l/m$, if $m\neq 0$ and $g(x)\neq 0$ anywhere on $\mathbf{I}$

(*e*) $\lim\limits_{x\to c}\,|f(x)|=|l|$

9. *Some important limits*

$$\lim_{x\to c}\frac{x^n-c^n}{x-c}=nc^{n-1}\ (c>0,\ n \text{ is a rational number})$$

$$\lim_{x\to 0}\frac{\sin x}{x}=1$$

$$\lim_{x \to 0} e^x = 1$$

$$\lim_{x \to 0} \frac{e^x - 1}{x} = 1,$$

$$\lim_{x \to 0} \frac{ln(1+x)}{x} = 1.$$

10. All polynomial functions, trigonometric functions, inverse trigonometric functions, the exponential function, and the logarithmic function are continuous at all points of their respective domains.

11. *Three important theorems for continuous functions* :

(*a*) Every function defined and continuous on a closed and bounded interval I is bounded therein.

(*b*) Every function defined and continuous on a closed and bounded interval attains its supremum and infimum.

(*c*) If f is continuous on the closed and bounded interval $[a, b]$, and c be any real number between $f(a)$ and $f(b)$, then there exists a real number x_0 in $]a, b[$ such that $g(x_0) = c$.

HISTORICAL NOTE

The Babylonians (C. 2000 B. C.) might be credited with a working definition of a function because of the tables prepared by them like the one for n^3+n^2, $n=1, 2, 3, \cdots, 30$ suggesting that a function is a table or correspondence between n in the left column and n^3+n^2 in the right column.

Rene Descartes (1637) might have been the first one to use the term function. He defined a function to mean any positive integral function of x.

Gottfried Wilhelm Leibnitz (1692) thought of a function as any quantity associated with a curve, such as the co-ordinates of a point on a curve, the length of the tangent to a curve etc.

In 1718, John Bernoulli, a student of Leibnitz gave for the first time a definition of a function free of geometric langauge. According to him *a function of a variable quantity is a magnitude formed in some manner from this variable quantity and constants.*

The next step in the development of the notation of a function is linked with the name of the famous swiss mathematician Leonhard Euler (1707-1783), a brilliant student of John Bernoulli. In his *Differential Calculus* he defined a function thus : *Quantities dependent on others such that as the second changes, so does the first are said to be functions.*

However, Euler and other mathematicians of his time required that a function must be defined by means of a formula. From the point of view of eighteenth century mathematicians the expression

$$y=\begin{cases} x, & \text{if } x<0 \\ x^2, & \text{if } x\geqslant 0 \end{cases}$$

defines not one but two functions.

It was J.B. Joseph Fourier (1768-1830) who in 1824 gave a new definition of a function, stressing that the main thing was the assignment of values for the function, whether the assignment was carried out by means of a single formula or not was unimportant. A few years latter (in 1837) Dirichlet (1805-1859) gave the modern definition of function. According to him *a variable quantity* y *is said to be a function* of a variable quantity x if to each value of the quantity x there corresponds a uniquely determined value of the quantity y. The definition of a function that we use today is essentially the same as given by Dirichlet.

□ □

SIR ISAAC NEWTON (1642-1727)

Isaac Newton was born in Woolsthorpe on Christmas day, 1642. When he was still a child, he showed great skill and delight in devising clever mechanical models and in conducting experiments. At the age of 18, he entered Trinity College, Cambridge. At the age of 23, he proved the generalized binomial theorem and created his Method of fluxions which is known today as the differential calculus. His method of Fluxions written in 1671 was not published until 1736. During the next one year, he performed his first experiments in optics, and formulated the basic principles of his theory of gravitation. In 1669 his teacher Barrow resigned the Lucasian professorship in his favour and Newton began his 18 years of University lecturing which he devoted to algebra and the theory of equations. In 1685, Newton completed the first book of his Principia, his greatest work which proved to be the most influential and the most admired work in the history of science. The complete treatise entitled Philosophiae naturalis principia mathematica was published in the middle of 1687 which immediately made an enormous impression throughout Europe. In 1689, Newton represented the University in parliament. In 1703, he was elected President of the Royal Society, a position to which he was annually re-elected until his death in 1727 after a lingering and painful illness and was buried in Westminster Abbey. Newton was a skilled experimentalist and a superb analyst. As a mathematician he is ranked almost universally as the greatest the world has ever produced. Newton was the greatest genius that ever lived. His accomplishments were poetically expressed by Pope in the lines :

Nature and Nature's laws lay, hid in night.
God said, 'Let Newton be,' and all was light.

CHAPTER 3

Derivatives

3·1. INTRODUCTION

The most remarkable achievement of the seventeenth century was the invention of the calculus towards the end of the century by Newton and Leibnitz. To some extent, it was an attempt to answer problems already tackled by the Greeks, but primarily the calculus was created to treat major scientific problems of the seventeenth century.

In the present chapter we shall introduce the notion of the derivative which is the basic theme of differential calculus. In the next chapter we shall study some simple applications of the derivative. In the fifth chapter we shall study the notion of the primitive (or anti-derivative) of a function which is the inverse of the notion of a derivative. We shall also see how it leads to a solution of the problem of determining the area bounded by curves.

3·2. RATE OF CHANGE

If you are sitting in a car that is running along a road, your eyes rest at the speedometer quite often and you keep on commenting in some such vein : *we are going at a speed of 75 km/h.* The needle in the speedometer turns a little this way and that way indicating the speed of the car at that particular instant. Are you wondering as to whatever can we mean by *the speed of the car at a particular instant* ? Let us denote time by t. Suppose the car starts at time $t=0$. Suppose the car travels a distance s metres in t seconds. There are two possibilities :

(*i*) The car is travelling at a constant or steady speed. In other words, its speed remains the same throughout its motion. Thus the car covers a distance s/t metres in one second This is the speed of the car. We describe this by saying that at every instant, the car is moving with a speed $(s/t)m/sec$ (s/t metres per second).

(*ii*) The speed of the car keeps on changing. Suppose it is changing according to the formula $s=t^2$ so that the distance covered at the end of 1, 2, 3,.........seconds is 1, 4, 9, ...metres respectively. Thus during the time $t=0$ and $t=1$ second, the car travels a distance 1 metre. During the time-interval from $t=1$ to $t=2$ seconds, it travels a

distance $4-1=3$ metres. The average speed of the car during this interval is, therefore,

$$\frac{\text{change in distance (in metres) during the time } t=1 \text{ to } t=2 \text{ (in seconds)}}{\text{change in time (in seconds)}}$$

$$=\frac{4-1}{2-1}=3.$$

Observe that the distance s depends upon time t. So s is a function of t, say $s=f(t)$, where $f(t)=t^2$. The average speed of the car during any time-interval can be calculated by the above formula. We shall now find the average speed of the car at $t=1$ second. For this we determine the average speeds of the car in short time intervals around 1 second. To be precise, we find the average speeds of the car in the interval $t=1$ to $t=1+h$ for some small values of h (positive as well as negative, but not zero) by the following formula :

$$\frac{\text{Change in distance (in metres) during time-interval from } t=1 \text{ to } t=1+h \text{ (in seconds)}}{\text{change in time (in seconds)}}$$

$$=\frac{f(1+h)-f(1)}{(1+h)-1}, \qquad (\because \; s=f(t))$$

$$=\frac{(1+h)^2-1^2}{h}, \qquad (\because \; f(t)=t^2)$$

$$=2+h. \qquad (\because \; h\neq 0)$$

Thus, when $h=0{\cdot}05$, $\dfrac{f(1+h)-f(1)}{(1+h)-1}=2+h=2+0{\cdot}05=2{\cdot}05.$

In this way, we get the following table :

h	$\{f(1+h)-f(1)\}/h$
·05	2·05
·04	2·04
·03	2·03
·02	2·02
·01	2·01
−·01	1·99
−·02	1·98
−·03	1·97
−·04	1·96
−·05	1·95

We have avoided the value $h=0$ because we cannot divide by 0. Now look at the above table carefully. The average speed $\dfrac{f(1+h)-f(1)}{h}$ over the time-interval from $t=1$ to $t=1+h$ approaches 2. In fact

$$\lim_{h\to 0}\frac{f(1+h)-f(1)}{h}=\lim_{h\to 0}\frac{(1+h)^2-1}{h}=2.$$

This time limit obviously gives us the average speed of the car at time $t=1$ because when h is very small the average speed during the interval from $t=1$ to $t=1+h$, more or less describes the speed at $t=1$. Thus using the concept of limit, we may not find the expression 'speed of the car at a given instant' so ridiculous. It is all a question of understanding. The more we know, the less we wonder :

Example 1. *Find the speed of the car at time $t=15$ in the above case.*

Solution. The speed of the car at $t=15$ is given by

$$\lim_{h\to 0}\frac{f(15+h)-f(15)}{h}=\lim_{h\to 0}\frac{(15+h)^2-15^2}{h}$$

$$=\lim_{h\to 0}\frac{(15+h+15)(15+h-15)}{h}$$

$$=\lim_{h\to 0}\frac{(30+h)(h)}{h},$$

$$=\lim_{h\to 0}(30+h), \qquad (\because h\neq 0)$$

$$=30.$$

Thus the speed at $t=15$ is 30 m/sec.

We shall as yet not get down the car that we had been riding above. We saw that a certain limit above gave us some idea of a certain rate of change viz., the rate at which the speed of the car is changing with change in time. We shall now probe into the geometrical meaning of the same limit. The function involved was $s=t^2=f(t)$. You know how to draw the graph of this function on s-axis and t-axis. It surely represents a parabola. We graph it as shown in Fig. 3·1.

For $t=1$ and $1+h$, $s=f(t)=f(1)$ and $f(1+h)$ respectively. This gives us two points P $(1, f(1))$, 0 $(1+h, f(1+h))$ lying on the parabola. Join PQ. What is the gradient of the line PQ ? You already know how to find out the gradient of a line through two given points. In fact,

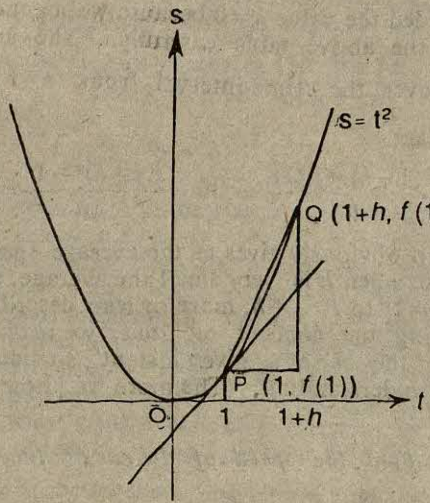


Fig. 3·1.

$$\text{grad. PQ} = \frac{f(1+h)-f(1)}{(1+h)-1},$$

$$= \frac{f(1+h)-f(1)}{h}.$$

As h tends to zero, the point Q approaches P on the parabola and the line PQ sort of just touches the parabola at P (*i.e.*, it becomes a tangent to the parabola at P). Thus $\lim\limits_{h \to 0} \dfrac{f(1+h)-f(1)}{h}$ expresses the gradient or slope of the tangent at the point $t=1$ of the curves $s=f(t)=t^2$.

Example 2. *Find the gradient of the tangent at the point $t=3$, to the curve $s=t^2$.*

Solution. The required gradient

$$= \lim_{h \to 0} \frac{f(3+h)-f(3)}{h}$$

$$= \lim_{h \to 0} \frac{(3+h)^2-3^2}{h},$$

$$= \lim_{h \to 0} \frac{(6+h) \cdot h}{h},$$

$$= \lim_{h \to 0} (6+h) = 6.$$

Example 3. *Find the rate of change of the volume with respect to the radius of a spherical balloon during the process of its being blown, at any one particular instant.*

Solution. Let us denote the volume at any instant by V. When the radius of the balloon is r cm, $V=\frac{4}{3}\pi r^3 \text{ cm}^3=f(r)$. As r changes, V also does. Let us find out the rate of change of V at the instant when $r=3$. This is given by

$$\lim_{h\to 0} \frac{f(3+h)-f(3)}{h} = \lim_{h\to 0} \frac{\frac{4}{3}\pi (3+h)^3-\frac{4}{3}\pi (3)^3}{h}$$

$$= \lim_{h\to 0} \tfrac{4}{3}\pi\left\{\frac{(3+h)^3-3^3}{h}\right\}$$

$$= \lim_{h\to 0} \tfrac{4}{3}\pi \left\{\frac{(3+h-3)\,\{(3+h)^2+3\ (3+h)+3^2\}}{h}\right\}$$

$$= \lim_{h\to 0} \tfrac{4}{3}\pi\ (h^2+9h+27),$$

$$= \lim_{h\to 0} \tfrac{4}{3}\pi \lim_{h\to 0}\ (h^2+9h+27),$$

$$= \tfrac{4}{3}\pi \{\lim_{h\to 0} h^2+ \lim_{h\to 0} 9h+ \lim_{h\to 0} 27\}$$

$$= \tfrac{4}{3}\pi \{(0+0+27)\},$$

$$= 36\pi.$$

Hence at $r=3$, the volume is changing atthe rate of $36\pi \text{ cm}^3/\text{cm}$.

In general, when the radius is r, the rate of change of the volume with respect to the radius is

$$\lim_{h\to 0} \frac{f(r+h)-f(r)}{h} = \lim_{h\to 0} \frac{\frac{4}{3}\pi\ (r+h)^3-\frac{4}{3}\pi\ r^3}{h},$$

$$= \lim_{h\to 0} \tfrac{4}{3}\pi\left\{\frac{(r+h)^3-r^3}{h}\right\},$$

$$= \lim_{h\to 0} \tfrac{4}{3}\pi \left\{\frac{h\ \{(r+h)^2+r\ (r+h)\ r^2\}}{h}\right\},$$

$$= \lim_{h\to 0} \tfrac{4}{3}\pi\ \{3r^2+3rh+h^2\},$$

$$= \tfrac{4}{3}\pi \lim_{h\to 0}\ (3r^2+3rh+h^2).$$

$$= \tfrac{4}{3}\pi\ .\ 3r^2=4\pi r^2.$$

So far we have considered rates of change of specific functions at specific points. Let us now consider an arbitrary function f and an arbitrary point a. We calculate the average rate of change in

the value of the function f with respect to the independent variable x in the interval from $x=a$ to $x=a+h$. This is given by the difference ratio

$$\frac{\text{Change in the value of the function}}{\text{Change in the variable}}$$
$$=\frac{f(a+h)-f(a)}{(a+h)-a}$$
$$=\frac{f(a+h)-f(a)}{h},\ (h\neq 0).$$

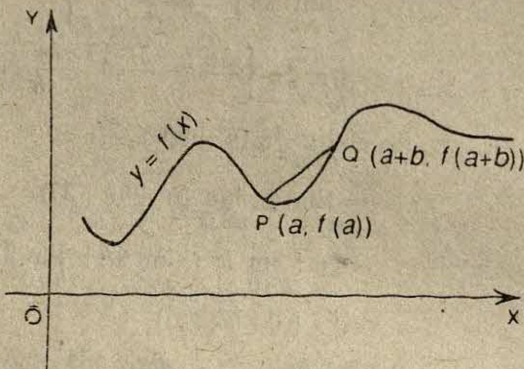


Fig. 3·2.

Geometrically, the above ratio represents the gradient of the chord PQ, where P and Q are points on the curve $y=f(x)$ with abscissae $x=a$ and $x=a+h$ respectively (see Fig. 3 2).

If $\lim\limits_{h\to 0}\dfrac{f(a+h)-f(a)}{h}$ exists, then it is called the rate of change of f at a. It is usual to denote this value by $f'(a)$.

Remark. As $h\to 0$, the point Q (see Fig. 3·2) approaches P. Also,

(*i*) the chord PQ tends to become the tangent to the curve $y=f(x)$ at the point P,

(*ii*) the gradient $\dfrac{f(a+h)-f(a)}{h}$ of PQ, tends to become the gradient of the tangent at P. Thus $\lim\limits_{h\to 0}\dfrac{f(a+h)-f(a)}{h}$ (in case it exists) is the gradient of the tangent to the curve $y=f(x)$ at the point $(a, f(a))$, and

(*iii*) $\frac{f(a+h)-f(a)}{h}$, the average rate of change in the interval $[a, a+h]$ tends to become the rate of change at a.

Example 4. *Find the average rate of change of the function f defined by*

$$f(x)=x+7, \forall x \in \mathbf{R}$$

in the interval $[2, 2+h]$ *and hence evaluate the rate of change of f at* $x=2$.

Solution. By definition, the average rate of change of f in the interval $[2, 2+h]$ is

$$\frac{f(2+h)-f(2)}{(2+h)-2}=\frac{\{(2+h)+7\}-(2+7)}{h},$$

$$=1.$$

Hence the rate of change of f at 2

$$=f'(2)$$

$$=\lim_{h\to 0}\frac{f(2+h)-f(2)}{h}$$

$$=\lim_{h\to 0} 1,$$

$$=1.$$

EXERCISE 3 (*a*)

1. Find the rate of change of the circumference of a circle with respect to the radius when the radius is 5 cm. (**Hint**. Find $f'(5)$, where $f(x)=2\pi x$).
2. Find the rate of change of the area of a circle with respect to the radius when the radius is 4 cm.
3. A ball is thrown vertically upwards. After t seconds of its being thrown, its height h is given by the formula $h=40t-6t^2$. Find its speed after 3 seconds of its being thrown.
4. The size n of a bacteria culture grows according to the formula $n=n_0+20t+3t^2$. Find the growth rate at $t=10$.
5. Complete the following table :

$f(x)$	a	$f'(a)$
$x-9$	2	
$2x^2+x+1$	7	
x^2+5	-1	
x^3	2	
6	7	

6. Find the average rate of change of functions given in the above problem over the interval $[5, 5+h]$ and find $f'(5)$ for each of them.

3·3. DERIVATIVE

In what follows we shall consider f to be a real-valued function whose domain is a subset I of **R**. For a point $a \in$ I, the rate of change of f at a is also called the derivative of f at a. Thus we have the following :

Definition 3·1. *Let $f: I \to \mathbf{R}$ be a function and let $x \in I$. If*

$$\lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$$

exists, then it is called the derivative of f at x and is denoted as $f'(x)$.

When the notation $y=f(x)$ is used, the derivative at x is also denoted as $\frac{dy}{dx}$, $\frac{d}{dx}[f(x)]$ or Df. The notation $\frac{dy}{dx}$ is due to Leibnitz. When f has a derivative at the point x (so that $f'(x)$ exists), we say that f is differentiable or derivable at x. If f is differentiable at each point of I, then f is said to be a differentiable function and the process of obtaining the derivatives (at points of I) is called differentiation. The function f' which as igns to each point x of I, the derivative $f'(x)$ at x is called the derived function of f.

Remark. Recall that the derivative is the limiting value of the difference ratio

$$\frac{f(x+h)-f(x)}{(x+h)-x} = \frac{\text{change in the value of } f}{\text{change in the variable } x},$$

but the change in the value of f depends upon the change in the value of x. To emphasize the fact that we are dealing with a change in x, it is usual to use the symbol Δx (read as delta x, Δ being a Greek capital letter pronounced 'delta') in place of h. The corresponding change in the value of f is then denoted by f. Thus

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x}.$$

3·3·1. The physical meaning of derivative

As we have seen in the section on rate of change, the rate of change of distance travelled by a particle with respect to time gives the velocity of the particle. In the next chapter, when we shall discuss some applications of the derivative, we shall find that the rate of change of velocity of a particle with respect to time gives the acceleration of the particle.

3·3·2. Geometrical meaning of the derivative

In the Section 3.1 we have seen that if $s=f(t)$ represents a

curve, then $\frac{ds}{dt}$ represents the slope of the tangent. In other words, if the equation of a curve be $y=F(x)$, then $F'(x)$ represents the slope of the tangent to the curve at the point 'x'. In the next chapter we shall discuss in detail as to how we can find the equation of the tangent to the curve at a given point.

Example 5. *Show that every constant function is differentiable. Also obtain the derivative at an arbitrary point.*

Solution. Let $f: x \to c$ be a constant function. Then

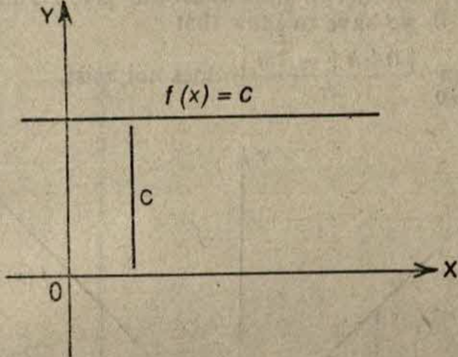


Fig 3·3.

$$\lim_{x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x},$$

$$= \lim_{\Delta x \to 0} \frac{c-c}{\Delta x}, \quad (\because \; f(x)=c \text{ for all } x \in I)$$

$$= \lim_{\Delta x \to 0} 0 = 0.$$

Hence f is differentiable and $f'(x)=0$, for all x in its domain. Geometrically, the tangent at every point on the curve has zero gradient (and so, is parallel to the x-axis) and is thus the line $f(x)=c$ itself.

Example 6. *Show that the function $f: x \to x+1$ is a differentiable function. Also, obtain the derived function of f.*

Solution. For any x,

$$\lim_{h \to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h \to 0} \frac{(x+h+1)-(x+1)}{h}$$

$$= \lim_{h \to 0} \frac{h}{h},$$

$$= \lim_{h \to 0} 1,$$
$$= 1.$$

Thus f is derivable at every point x of its domain. This proves that f is a differentiable function. Since $f'(x)=1$ for all $x \in \mathbf{R}$, therefore the derived function f' is a constant function taking the value 1 at every point of $\mathbf{R}$.

Example 7. *Show that the function $f : x \to |x|$ is derivable at every point except zero.*

Solution. In order to show that the given function is not derivable at $x=0$, we have to show that

$$\lim_{h \to 0} \frac{|0+h| - |0|}{h} \text{ does not exist.}$$

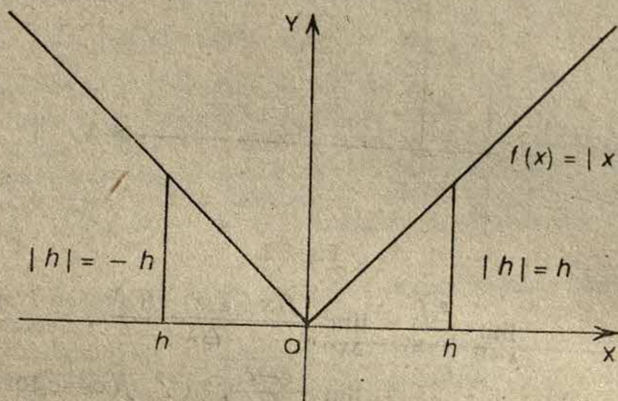


Fig. 3·4.

or equivalently that $\lim_{h \to 0} \frac{|h|}{h}$ does not exist.

Now, let us see how does $|h|/h$ behave in any neighbourhood of 0. In other words let us consider values of h quite close to 0 and see whether the values of $|h|/h$ can be made as close as we please to a number l. For h small but $h>0$, $|h|/h=h/h=1$. For h small but $h<0$, $|h|=-h$ so that $|h|/h=-h/h=-1$. Thus there cannot exist *any* number l such that $\left|\frac{|h|}{h}-l\right|$ would be arbitrarily small in any neighbourhood of zero. Hence we conclude that the given function is not derivable at zero.

Let now $a>0$. Choose h so that $|h|<a$. Then whether $h>0$ or $h<0$, $a+h>0$, so that

$$\lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$$

$$= \lim_{h\to 0} \frac{|a+h|-|a|}{h},$$

$$= \lim_{h\to 0} \frac{(a+h)-a}{h}$$

$$= \lim_{h\to 0} 1,$$

$$= 1.$$

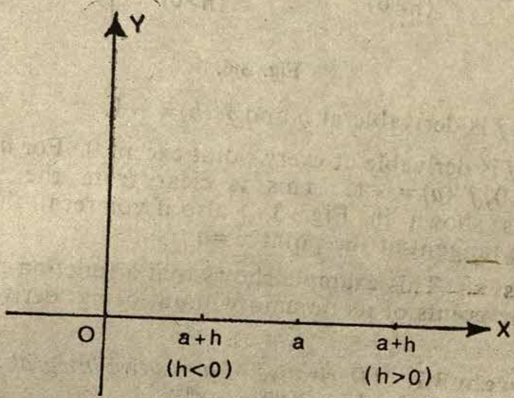


Fig. 3·5.

Hence f is derivable at a. Moreover, $f'(a)=1$.

Consider now a negative real number b. Choose h so that $|h| < |b|$. Then whether $h>0$ or $h<0$, $b+h<0$. Hence

$$\lim_{h\to 0} \frac{f(b+h)-f(b)}{h}$$

$$= \lim_{h\to 0} \frac{|b+h|-|b|}{h}$$

$$= \lim_{h\to 0} \frac{-(b+h)-(-b)}{h}$$

$$= \lim_{h\to 0} \frac{-h}{h}.$$

$$= \lim_{h\to 0} (-1),$$

$$= -1.$$

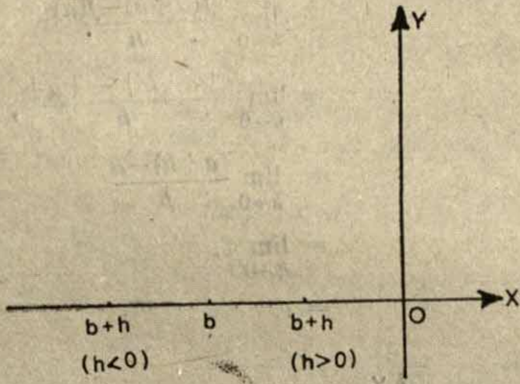


Fig. 3·6.

Hence f is derivable at b and $f'(b)=-1$.

Thus f is derivable at every point except 0. For $a>0, f'(a)=1$ and for $a<0, f'(a)=-1$. This is clear from the graph of the function (as shown in Fig. 3·4) also if you recall that $f'(a)$ is the slope of the tangent at the point $x=a$.

Remark. This example shows that a function may be derivable at some points of its domain without being derivable at some others.

Theorem 3·1. (*Derivative of a positive integral power of* x) *If* n *is a positive integer, then* $D(x^n)=nx^{n-1}$.

Proof. In order to obtain $D(x^n)$, we have to determine the

limit $\quad \lim\limits_{h\to 0}\dfrac{(x+h)^n-x^n}{h}$ (in case it exists)

Now $\lim\limits_{h\to 0}\dfrac{(x+h)^n-x^n}{h}=\lim\limits_{h\to 0}\dfrac{(x^n+nhx^{n-1}+\dots+h^n)-x^n}{h}$,

by the binomial theorem,

$$=\lim_{h\to 0}\left\{nx^{n-1}+\frac{n(n-1)}{2}hx^{n-2}+\dots+h^{n+1}\right\},$$

$$=\lim_{h\to 0}\{nx^{n-1}+hg(x, h)\},$$

where $g(x, h)$ is a polynomial in x and

$$=\lim_{h\to 0}(nx^{n-1})+\lim_{h\to 0}(h)\,.\lim_{h\to 0}g(x, h),$$

$$=nx^{n-1}+0\,.\lim_{h\to 0}g(x, h),$$

$$=nx^{n-1}.$$

Thus $$\boxed{Dx^n = nx^{n-1}}$$

Example 8. *Show that the function f defined by $f(x)=\sqrt{x}$, $x>0$, is differentiable.*

Solution.
$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$$
$$=\lim_{h\to 0}\frac{\sqrt{x+h}-\sqrt{x}}{h},$$
$$=\lim_{h\to 0}\frac{(\sqrt{x+h}-\sqrt{x})(\sqrt{x+h}+\sqrt{x})}{h(\sqrt{x+h}+\sqrt{x})},$$
$$=\lim_{h\to 0}\frac{(x+h)-x}{h(\sqrt{(x+h)}+\sqrt{x})},$$
$$=\lim_{h\to 0}\frac{1}{\sqrt{(x+h)}+\sqrt{x}}$$
$$=\frac{1}{2\sqrt{x}} \quad \text{because } \lim_{h\to 0}\sqrt{x+h}=\sqrt{x}.$$

Remark. Later on we shall obtain the derivative of an arbitrary rational power of x, whenever it exists.

EXERCISE 3 (*b*)

Show that each of the functions in problems 1—5 below is derivable at $x=1$. Also, find $f'(1)$ in each case.

1. $f: x \to 10$.
2. $f: x \to x$.
3. $f: x \to 2x$.
4. $f: x \to 2x+41$.
5. $f: x \to ax+b$, where a and b are fixed real numbers.

Find $\frac{dy}{dx}$ for each of the functions in problems 6—13 :

6. $y=x^2$.
7. $y=x^2+3$.
8. $y=2x^2+5$.
9. $y=-x^2+1$.
10. $y=|x+1|$.
11. $y=|x|+5$.
12. $y=-|x|$.
13. $y=2|x|$.
14. Which of the functions in problems 1—13 above are differentiable ?
15. For each of the non-differentiable functions in problems 1—13 above, find the points at which it is not differentiable.
16. Find $D'[f(x)]$ for each of the following :

(*a*) $f(x)=x^5$ (*b*) $f(x)=x^{10}$

(*c*) $f(x)=x^{100}$ (*d*) $f(x)=x^{10^3}$

(e) $f(x)=\sqrt{(x+2)}$ (f) $f(x)=\sqrt{5x}$

(g) $f(x)=\sqrt{3x+7}$.

3·4. ALGEBRA OF DERIVATIVES

Before we proceed with the task of finding derivatives of some standard functions, we shall develop an algebra of the derivatives so as to facilitate later calculations.

3·4·1. Derivative of a Scalar Multiple of a Function

Consider the function $y=cf(x)$, where c is a constant. The derivative of y with respect to x (if it exists !) is given by

$$\lim_{h\to 0} \frac{(cf)(x+h)-(cf)(x)}{h} = \lim_{h\to 0} \left\{ \frac{cf(x+h)-cf(x)}{h} \right\}$$

$$= \lim_{h\to 0} c \left\{ \frac{f(x+h)-f(x)}{h} \right\}$$

$$= \lim_{h\to 0} c . \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

$$= cf'(x) \text{ (provided } f'(x) \text{ exists !).}$$

This proves :

Theorem 3·2. *If f is a differentiable function, so is cf, and*

$$\boxed{\begin{array}{ll} & D((cf)(x))=cD(f(x)) \\ i.e., & (cf)'(x)=cf'(x) \end{array}}$$

3·4·2. Derivative of the Sum of two Differentiable Functions

Let f and g be differentiable functions from **R** to **R**. Let us examine whether $f+g$, the sum of the functions f and g is differentiable.

Now $$\lim_{h\to 0} \frac{(f+g)(x+h)-(f+g)(x)}{h}$$

$$= \lim_{h\to 0} \frac{\{f(x+h)+g(x+h)\}-\{f(x)+g(x)\}}{h}$$

$$= \lim_{h\to 0} \frac{\{f(x+h)-f(x)\}+\{g(x+h)-g(x)\}}{h}$$

$$= \lim_{h\to 0} \left[\frac{f(x+h)-f(x)}{h} + \frac{g(x+h)-g(x)}{h} \right]$$

$$= \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}\right] + \lim_{h\to 0}\left[\frac{g(x+h)-g(x)}{h}\right]$$

$$=f'(x)+g'(x).$$

This proves :

Theorem 3·3. *The derivative of the sum of two differentiable functions f and g is a differentiable function and its derived function is the sum of the derived functions of f and g, i.e ,*

$$\boxed{\begin{array}{l} D((f+g)(x))=D(f(x))+D'g(x)) \\ i.e., \quad (f+g)'(x)=f'(x)+g'(x) \end{array}}$$

Remark. The above result can be extended to a finite sum instead of the sum of two functions.

3·4·2. Derivative of the Product of Two Differentiable Functions

Let f and g be two differentiable functions from **R** to **R.** We want to find out whether their product fg is also differentiable.

Now $$\lim_{h\to 0}\frac{(fg)(x+h)-(fg)(x)}{h},$$

$$= \lim_{h\to 0}\frac{f(x+h)\,g(x+h)-f(x)\,g(x)}{h},$$

$$= \lim_{h\to 0}\frac{\{f(x+h)-f(x)\}g(x+h)+\{g(x+h)-g(x)\}f(x)}{h}$$

$$= \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}\,g(x+h)\right] + \lim_{h\to 0}\left[\frac{g(x+h)-g(x)}{h},\,f(x)\right],$$

$$= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}\lim_{h\to 0}g(x+h) + \lim_{h\to 0}\frac{g(x+h)-g(x)}{h}\lim_{h\to 0}f(x)$$

$$=f'(x)g(x)+g'(x)f(x).$$

Hence the following :

Theorem 3·4. *The product of two differentiable functions is again a differentiable function and its derivative at any point x is given by the formula*

$$\boxed{\begin{array}{ll} & D((fg(x))=D(f(x)).g(x)+f(x).D(g(x)) \\ \text{i.e.,} & (fg)'(x)=f'(x)g(x)+f(x)g'(x) \end{array}}$$

Corollary. *If $f_1, f_2, \ldots\ldots, f_n$ are differentiable functions, then* $(f_1 f_2 \ldots\ldots f_n)'(x)=f'(x)f_2(x)\ldots\ldots f_n(x)+f_1(x)f_2'(x)f_3(x)\ldots\ldots f_n(x)+\ldots\ldots$
$\ldots\ldots+f_1(x)f_2(x)\ldots\ldots f_n'(x)$.

3·4·5. Derivative of the Quotient of Two Differentiable Functions

Let $F=f/g$, where f and g are functions from **R** to **R** and $g(x)\neq 0$ for any x. Then

$$\lim_{h\to 0} \frac{F(x+h)-F(x)}{h}$$

$$=\lim_{h\to 0} \frac{(f/g)(x+h)-(f/g)(x)}{h}$$

$$=\lim_{h\to 0} \frac{\dfrac{f(x+h)}{g(x+h)}-\dfrac{f(x)}{g(x)}}{h}$$

$$=\lim_{h\to 0} \frac{g(x)f(x+h)-f(x)\,g(x+h)}{hg(x+h)g(h)}$$

$$=\lim_{h\to 0} \left[\frac{g(x)\{f(x+h)-f(x)\}-f(x)\,\{g(x+h)-g(x)\}}{hg(x+h)\,g(x)}\right]$$

$$=\lim_{h\to 0} \frac{\left[g(x)\left\{\dfrac{f(x+h)-f(x)}{h}\right\}-f(x)\left\{\dfrac{g(x+h)-g(x)}{h}\right\}\right]}{\lim\limits_{h\to 0}\{g(x+h)g(x)\}}$$

$$=\lim_{h\to 0} \frac{\left[g(x)\left\{\dfrac{f(x+h)-f(x)}{h}\right\}\right]-\lim\limits_{h\to 0}\left[f(x)\left\{\dfrac{g(x+h)-g(x)}{h}\right\}\right]}{\lim\limits_{h\to 0} g(x+h)\lim\limits_{h\to 0} g(x)}$$

$$=\frac{\lim\limits_{h\to 0} g(x)\lim\limits_{h\to 0}\dfrac{f(x+h)-f(x)}{h}-\lim\limits_{h\to 0} f(x)\lim\limits_{h\to 0}\dfrac{g(x+h)-g(x)}{h}}{g(x)\,g(x)}$$

$$=\frac{g(x)f'(x)-f(x)\,g'(x)}{(g(x))^2}$$

Hence the following :

Theorem 3·5. *The quotient f/g of two differentiable functions f and g such that $g(x)\neq 0$ for any x, is again a differentiable function and its derivative at any point x is given by the formula*

$$\boxed{\left(\frac{f}{g}\right)'(x)=\frac{g(x)f'(x)-f(x)g'(x)}{(g(x))^2}}$$

Theorem 3·6. (*Derivative of a Negative Power of x*).

$$D(x^n)=nx^{n-1},$$

where n is a negative integer and $x\neq 0$.

Proof. Consider the function $f: x\to x^{-m}$, where $m\in N$.

The function f may be regarded as the quotient g/g_1 of the functions g and g_1, where $g(x)=1$ for all $x\in \mathbf{R}$ and $g_1(x)=x^m$ for all $x\in \mathbf{R}$. Now g is differentiable, being a constant function. Also, g_1 is a differentiable function and $g_1(x)\neq 0$ if $x\neq 0$. So except at $x=0$, we find that the given function $f=g/g_1$ has a derivative.

Thus, if $x\neq 0$.

$$f'(x)=(g/g_1)'(x),$$

$$=\frac{g_1(x)g'(x)-g(x)g_1'(x)}{(g_1(x))^2} \text{ (by Theorem 3·4)}$$

$$=\frac{x^m \,.\, 0-1\,.\,mx^{m-1}}{(x^m)^2}, \text{ (derivative of a constant is zero),}$$

$$=-mx^{m-1}/x^{2m},$$

$$=-mx^{-m-1}.$$

Denoting $-m$ by by n, we have

$$f(x)=x^n \text{ (}n\text{, a negative integer) and } f'(x)=nx^{n-1}.$$

Example 9. *Differentiate the function* $f: x\to x^3+3x^2-7$ *and find the rate of change of f at* $x=5$.

Solution.
$$f'(x)=\frac{d}{dx}(x^3+3x^2-7),$$

$$=\frac{d}{dx}(x^3)+\frac{d}{dx}(3x^2)+\frac{d}{dx}(-7),$$

(by Theorem 3·2)

$$=3x^2+3\frac{d}{dx}(x^2)+0, \text{ (by Theorem 3·1)}$$

$$=3x^2+3(2x),$$

$$=3x^2+6x.$$

The rate of change at $x=5$ is nothing but $f'(5)$ and

$$f'(5)=3\cdot 5^2+6\cdot 5$$

$$=75+30=105.$$

Example 10. *Differentiate the function*

$$f : x \to (x^{-3}+1)\,(x^{10}-13).$$

Solution. $f'(x) = \dfrac{d}{dx}\left\{(x^{-3}+1)\,(x^{10}-13)\right\}$

$$=(x^{10}-13)\,\frac{d}{dx}(x^{-3}+1)+(x^{-3}+1)\frac{d}{dx}(x^{10}-13),$$

(by Theorem 3·3)

$$=(x^{10}-13)\left\{\frac{d}{dx}(x^{-3})+\frac{d}{dx}(1)\right\}$$

$$+(x^{-3}+1)\left\{\frac{d}{dx}(x^{10})+\frac{d}{dx}(-13)\right\}$$

$$=(x^{10}-13)(-3x^{-4}+0)+(x^{-3}+1)(10x^{9}+0)$$

$$=-3(x^{6}-13x^{-4})+10(x^{6}+x^{9})$$

$$=10x^{9}+7x^{6}+36x^{-4}.$$

EXERCISE 3 (*c*)

1. Prove that if f and g are differentiable functions, then
$$d(f(x)-g(x))=\mathrm{D}(f(x))-\mathrm{D}(g(x)).$$

2. Find $\dfrac{dy}{dx}$ for each of the following functions :

(*a*) $y=3x^2+4x+5$ (*b*) $y=3x^3-5x^2+9$

(*c*) $y=6x^4+3x^2+1$ (*d*) $y=5x^9+6x^7+x$

(*e*) $y=x^{15}-11x+1$ (*f*) $x=100x^{100}$

(*g*) $y=a_0+a_1x+a_2x^2+\cdots\cdots+a_nx^n$,

where $a_0, a_1, a_2, \cdots\cdots, a_n$ are fixed real numbers.

3. Find $f'(x)$ for each of the following functions f:

(*a*) $f : x \to x^{-3}+x+x^3$ (*b*) $f : x \to x^{-7}+x^{-5}+2.$

(*c*) $f : x \to 3x^{-1}+5+x^6$ (*d*) $f : x \to 5x^{-12}+x^{-4}+x$

(*e*) $f : x \to (x+x^{-1})^2.$ (*f*) $f : x \to (x^{-2}+2)^3.$

4. Find $\mathrm{D}(f(x))$ for each of the following functions f :

(*a*) $f(x)=\dfrac{2x+3}{5x+7}$ (*b*) $f(x)=\dfrac{5x^2+3x+1}{3x+10}$

(*c*) $f(x)=\dfrac{x^3+1}{x^2-1}$ (*d*) $f(x)=\dfrac{x^{-3}+x^{-1}}{x+2x^3}$

(*e*) $f(x)=\dfrac{1+x^{-1}}{x^{-2}+1}$ (*f*) $f(x)=\dfrac{9x^2+2x^{-10}}{2x+5x^{-7}}$

(g) $f(x)=\dfrac{a+bx+cx^2}{x+qx^{-1}+rx^{-2}}$

where a, b, c, p, q, r are fixed real numbers.

5. Differentiate each of the following and obtain $f'(2)$ for each :

(a) $f(x)=\{(x^2+1)^2+3\}^2$ (b) $f(x)=\{(x^3+1)^2+5\}^2$

(c) $f(x)=\{(x^2+1)^2-5\}^3$ (d) $f(x)=\{(2x^3-1)^3-3\}^3$.

3·5. CHAIN RULE OF DIFFERENTIATION

You can now differentiate any polynomial function. This enables you to differentiate functions such as

$$\{(x^2+3x)^3+5\}^{10}$$

even though it may require a lot of time and patience as your experience with problems like 5(c) and 5(d) above will tell you. It is said that we owe a lot many inventions to people who were lazy enough not to perform certain time/labour/patience consuming operations but were intelligent enough to find a way out. Who knows if the chain-rule or differntiation (using which, differentiating functions such as above may be a matter of not more than a few seconds) is not a brain child of such a person. In any case, let us now learn this delightful rule.

Theorem 3·6. *Let $y=g(u)$ and $u=f(x)$. If both $\dfrac{dy}{du}$ and $\dfrac{du}{dx}$ exist, then $\dfrac{dy}{dx}$ exists and is given by*

$$\frac{dy}{dx}=\frac{dy}{du}\cdot\frac{du}{dx}.$$

Proof. Note first of all that

$$y=g(u)=g(f(x))=(g\circ f)(x),$$

so that y is the composite function $g\circ f$. We are given that y, regarded as a function of u, is differentiable. We wish to prove that y regarded as a function of x is also differentiable. To do this (following our notation in the remark on page 158) we have to show that

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}$$

exists.

Now Δu, being the change in the value of u corresponding to a change Δx in the value of x, is given by

$$\Delta u=f(x+\Delta x)-f(x).$$

Also, f being a differentiable function of x,

$$\frac{du}{dx}=f'(x)=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}=\lim_{x\to 0}\frac{\Delta u}{\Delta x},$$

so that

$$\lim_{\Delta x\to 0}\Delta u=\lim_{\Delta x\to 0}\left\{\frac{\Delta u}{\Delta x}.\,\Delta x\right\},$$

$$=\lim_{\Delta x\to 0}\left\{\frac{\Delta u}{\Delta x}\right\}\lim_{\Delta x\to 0}\{\Delta x\},$$

$$=\frac{du}{dx}\,.\,0,$$

$$=0.$$

This means that

$$\Delta x\to 0\Rightarrow\Delta u\to 0. \qquad ...(1)$$

We assume, however, that $\Delta u\neq 0$.

Now

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\left\{\frac{\Delta y}{\Delta u}\cdot\frac{\Delta u}{\Delta x}\right\},$$

$$=\lim_{\Delta x\to 0}\left\{\frac{\Delta y}{\Delta x}\right\}.\lim_{\Delta x\to 0}\left\{\frac{\Delta u}{\Delta x}\right\}$$

$$=\lim_{\Delta x\to 0}\left\{\frac{\Delta y}{\Delta x}\right\}\cdot\frac{du}{dx},\text{ using (1)},$$

$$=\frac{dy}{du}\cdot\frac{du}{dx}$$

Hence $\frac{dy}{dx}$ exists and is equal to $\frac{dy}{du}\cdot\frac{du}{dx}$.

Remark. Put differently, the above theorem says that if $h(=g\circ f)$ is the composite of two differentiable functions g and f, then h is differentiable and $h'(x)=g'(f(x))\,.\,f'(x)$.

Example 11. *Differentiate* $(x^2+1)^3$ *with respect to* x.

Solution. Let $u=x^2+1$. Then $y=u^3$. Since y is a differentiable function of u and u is a differentiable function of x, therefore using chain rule

$$\frac{dy}{dx}=\frac{dy}{du}\cdot\frac{du}{dx},$$

$$=\frac{d}{du}(u^3)\,.\,\frac{d}{dx}(x^2+1),$$

$$(\because\ y=u^3\text{ and }u=x^2+1)$$

$$=3u^2 \cdot 2x,$$
$$=6x(x^2+1)^2$$

Remark. Check the solution by expanding $(x^2+1)^3$ using the binomial theorem and then differentiating.

Example 12. *Find* $\frac{dy}{dx}$ *when* $y=(2x^3+1)^{501}$.

Solution. Let $u=2x^3+1$, so that $y=u^{501}$. Since $\frac{dy}{du}$ and $\frac{du}{dx}$ both exist, therefore by the chain rule,

$$\frac{dy}{dx}=\frac{dy}{du}\cdot\frac{du}{dx},$$
$$=501u^{500}\cdot 6x^2,$$
$$=3006x^2(2x^3+1)^{500}$$

Remarks 1. *Instead of introducing u explicitly every time while applying the chain rule, after a little practice you would find it more convenient to do away with u and arrange the working in the above example as follows :*

$$\frac{dy}{dx}=\frac{d}{d(2x^3+1)}[(2x^3+1)^{501}]\cdot\frac{d}{dx}(2x^3+1).$$
$$=501\,(2x^3+1)^{500}\cdot 6x^2, \text{ etc.}$$

2. Would you dare solving the above example without using the chain rule ?

Example 13. *Differentiate the function* $h: x \to (x^3+7x^2)^{100}$.

Solution. $h=g\circ f$ where $f: x\to x^3+7x^2$, $g: x\to x^{100}$.

Thus $$f'(x)=\frac{d}{dx}(x^3+7x^2),$$
$$=\frac{d}{dx}(x^3)+\frac{d}{dx}(7x^2),$$
$$=3x^2+14x. \qquad \ldots(1)$$

$g'(f(x))$ means the derivative of g with respect to $f(x)$.

Writing $$f(x)=x^3+7x^2=y,$$
$$g(f(x))=g(y)=y^{100},$$
and $$g'(f(x))=g'(y),$$
$$=100y^{99},$$
$$=100(x^3+7x^2)^{99}. \qquad \ldots(2)$$

Thus $$h'(x)=(g\circ f)'(x),$$
$$=g'(f(x))\cdot f'(x),$$
$$=100(x^3+7x^2)^{99}(3x^2+14),$$
from (1) and (2).

Example 14. *Find* $\dfrac{dy}{dx}$, *where* $y=\{(2x+3)^2+5\}^3$.

Solution. Let $v=2x+3$ and $u=(2x+3)^2+5$, so that $y=u^3$, $u=v^2+5$.

Now $\dfrac{dy}{dx}$ and $\dfrac{du}{dx}$ both exist, so that by the chain rule,

$$\frac{dy}{dx}=\frac{dy}{du}\cdot\frac{du}{dx}. \qquad \text{...(1)}$$

Again, $\dfrac{du}{dv}$ and $\dfrac{dv}{dx}$ both exist, therefore

$$\frac{du}{dx}=\frac{du}{dv}\cdot\frac{dv}{dx}. \qquad \text{...(2)}$$

From (1) and (2),

$$\begin{aligned}\frac{dy}{dx}&=\frac{dy}{du}\ \frac{du}{dv}\cdot\frac{dv}{dx},\\ &=\frac{d}{du}(u^3)\cdot\frac{d}{dv}(v^2+5)\cdot\frac{d}{dx}(2x+3),\\ &=3u^2\cdot 2v\cdot 2,\\ &=12(2x+3)\{(2x+3)^2+5\}^2.\end{aligned}$$

Remark. The above example illustrates a situation where more than one application of the chain rule may be desirable.

EXERCISE 3 (*d*)

Find $\dfrac{dy}{dx}$ for each of the following:

1. $y=(2x+11)^7$.
2. $y=(3x-7)^{52}$.
3. $y=(7x+6)^2+9$.
4. $y=(-2x+1)^{53}-5$.
5. $y=\dfrac{3}{(1+2x+x^2)}$.
6. $y=\dfrac{1}{(2x+3x^2+6x^3)^2}$.
7. $y=(2x+3)^2+(x+13)^3$.
8. $y=(x+5)^{-7}+(9x+1)^{11}$.
9. $y=\dfrac{(2x+3)^2}{(7x+9)^3}$.
10. $y=\dfrac{(3+5x^2+7x^3)^9}{(9+11x^2+13x^3)^{11}}$.
11. $y=\{(x^2+x+1)^2+3\}^2$.
12. $y=\{(2x^3-3)^3+x\}^{10}$.
13. $y=\{x+(x+1)^2+(x^2+x+1)^2\}^2$.
14. $y=\{(2x+5)^3+(9x-5)^{-3}\}^7$.
15. $y=[\{(1+x^2)^2+x^4\}^3+x^8]^{16}$.

3·6. DERIVATIVES OF INVERSE FUNCTIONS

Let us recall that two functions f and g are said to be inverses of each other if $g\circ f$ and $f\circ g$ are the identity functions, *i.e.*, if $g(f(x))=x$ for all $x\in dom\, f$, and $f(g(y))=y$ for all $y\in dom\, g$.

If we are given two functions f and g which are inverses of each other, and if we know that both possess derivatives, than we can apply the chain rule for differentiation. We shall then have

$$g'(f(x))\, f'(x)=\frac{d}{dx}\,(x)=1.$$

Similarly $f'(g(y))k'(y)=1$.

From the above relations we find that both the functions f and g can be differentiable only if neither of the derivatives is zero. It can be proved that conversely, if one of the functions has a derivative that is not zero, then the other also has a non-zero derivative. This we state more precisely in the following :

Theorem 3 7. (*Inverse Function Theorm for derivatives*). *Let $y=f(x)$ and $x=g(y)$ be strictly monatone functions which are inverses of each other. If f has at x_0 the derivative $f'(x_0)\neq 0$, then g has at $y_0(=f'(x_0))$ the derivative*

$$g'(y_0)=\frac{1}{f'(x_0)}.$$

Remark. In Leibnitz notation, the above rule would look thus :

$$\frac{dx}{dy}=\frac{1}{\frac{dy}{dx}}$$

3·7. DERIVATIVES OF RADICAL FUNCTIONS

So far we have obtained the derivative of x^n where n is an integer. We can now obtain, by applying the inverse function theorem, the derivative of x, where r is any rational number.

Theorem 3·8. *Let r be a rational number and let $y=f(x)=x^r$. Then*

$$\frac{dy}{dx}=r\,x^{r-1},$$

for all those x for which x^{r-1} and x^r are defined.

Proof. If r is an integer, then the result holds by Theorem 3·1 and Theorem 3.6.

Let us now consider the case when $r=1/q$, q being a positive integer. In order to prove the result in this case, we consider the functions

$$y=f(x)=x^{1/q}, \text{ and } x=g(y)=y^q.$$

The functions f and g are inverses of each other, and therefore, by the inverse function theorem

$$\frac{dy}{dx}=\frac{1}{\frac{dx}{dy}},$$

$$=\frac{1}{qy^{q-1}}$$

$$=\frac{1}{qx^{(q-1)/q}}$$

$$=(1/q)x^{(1/q)-1} \qquad ...(1)$$

Let us now assume that $r=p/q$, where p is an integer, q is a positive integer.

Then $\frac{d}{dx}(x^r)=\frac{d}{dx}\{[x^{1/q}]^p\}$

$$=p(x^{1/q})^{p-1}.\frac{d}{dx}(x^{1/q}), \text{ by the chain rule,}$$

$$=p\,(x^{1/q})^{p-1}.\ \frac{1}{q}x^{(1/q)-1}, \text{ by (1)}$$

$$=(p/q)x^{(p/q)-1},$$

$$=rx^{r-1}.$$

Remark. The above rule enables us to differentiate radical functions.

Example 15. *Differentiate* $(1+x^2)^{5/7}$.

Solution. We have

$$\frac{d}{dx}[(1+x^2)^{5/7}]$$

$$=\frac{d}{d(1+x^2)}[(1+x^2)^{5/7}]\cdot\frac{d}{dx}[1+x^2], \qquad \text{by chain rule}$$

$$=\frac{5}{7}(1+x^2)^{(5/7)-1}.\,2x,$$

$$=\frac{10}{7}x\,(1+x^2)^{-2/7}.$$

Example 16. *Differentiate*

$$f(x)=\sqrt[3]{[\sqrt[5]{x}+\sqrt[9]{x}\,]}.$$

Solution. $f'(x)=\frac{d}{dx}[(x^{1/5}+x^{1/9})^{1/3}]$,

$$=\frac{d[(x^{1/5}+x^{1/9})^{1/3}]}{d\,(x^{1/5}+x^{1/9})}\ \frac{d\,(x^{1/5}+x^{1/9})}{dx}$$

$$=\frac{d}{d(x^{1/5}+x^{1/9})}[(x^{1/5}+x^{1/9})^{1/3}.\ \frac{d}{dx}[x^{1/5}+x^{1/9}]$$

$$=\tfrac{1}{3}(x^{1/5}+x^{1/9})^{-2/3}.[\tfrac{1}{5}x^{-4/5}+\tfrac{1}{9}x^{-8/9}].$$

EXERCISE 3 (*e*)

Differentiate each of the following :

1. $(2x+3)^{1/3}$.
2. $(3x-5)^{5/7}$.
3. $(3x^2+1)^{2/3}$.
4. $(2x^2-5x+3)^{1/6}$.
5. $(2x+3)^{1/5}+(x-2)^{1/3}$.
6. $x^{2/3}(x^2+1)^{1/2}$.
7. $x^{3/4}(2x-1)^{1/2}$.
8. $(3x-2)^{1/2}(x+4)^{1/3}$.
9. $(x^{1/2}+x^{1/3})^2$.
10. $(x^{2/3}-x^{3/4})^{1/2}$.

12. $(4x+1)^{1/2}+(x^2-2)^{2/3}$.

12. $\sqrt{2x^2+1}+3\sqrt[3]{4-x}$.

3·8. DIFFERENTIATION OF TRIGONOMETRIC FUNCTIONS

We shall now obtain the derivatives of the trigonometric functions $\sin x$, $\cos x$, etc. However, while doing so we shall need the following limit theorem :

$$\lim_{x\to 0}\frac{\sin x}{x}=1.$$

3·8·1. Derivative of sin x

If $\qquad y=f(x)=\sin x,$

then $\qquad f(x+h)=\sin(x+h),$

and
$$\frac{dy}{dx}=\lim_{h\to 0}\frac{\sin(x+h)-\sin x}{h}$$
$$=\lim_{h\to 0}\frac{2\sin(\tfrac{1}{2}h)\cos(x+\tfrac{1}{2}h)}{h}$$
$$=\lim_{h\to 0}\left\{\frac{\sin(\tfrac{1}{2}h)}{\tfrac{1}{2}h}\cos(x+\tfrac{1}{2}h)\right\}$$
$$=\cos x,$$

since $\displaystyle\lim_{h\to 0}\frac{\sin(\tfrac{1}{2}h)}{\tfrac{1}{2}h}=1,\ \lim_{h\to 0}\cos(x+\tfrac{1}{2}h)=\cos x.$

Hence

$$\boxed{D(\sin x)=\cos x}$$

3·8·2. Derivative of cos x

If $\qquad y=f(x)=\cos x,$

then $\qquad f(x+h)=\cos(x+h),$

and
$$\frac{dy}{dx}=\lim_{h\to 0}\frac{\cos(x+h)-\cos x}{h},$$

$$=\lim_{h\to 0}\frac{-2\sin(\frac{1}{2}h)\sin(x+\frac{1}{2}h)}{h}$$

$$=\lim_{h\to 0}\left\{(-1)\frac{\sin(\frac{1}{2}h)}{\frac{1}{2}h}\sin(x+\tfrac{1}{2}h)\right\}$$

$$=-\sin x,$$

since $\lim_{h\to 0}\frac{\sin(\frac{1}{2}h)}{\frac{1}{2}h}=1$, $\lim_{h\to 0}\sin(x+\frac{1}{2}h)=\sin x$.

Hence

$$\boxed{D(\cos x)=-\sin x}$$

Remark. We could have also proceeded thus :

Since $\cos x=\sin(x+\pi/2)$, therefore,

$$D(\cos x)=\cos(x+\pi/2).\ D(x+\pi/2),$$
$$=(-\sin x).\ 1,$$
$$=-\sin x.$$

3·8·3. Derivative of tan x

If $\quad y=f(x)=\tan x,$

then $\quad f(x+h)=\tan(x+h),$

and $\quad \frac{dy}{dx}=\lim_{h\to 0}\frac{\tan(x+h)-\tan x}{h},$

$$=\lim_{h\to 0}\frac{\sin(x+h)\cos x-\cos(x+h)\sin x}{h\cos x\cos(x+h)},$$

$$=\lim_{h\to 0}\left\{\frac{\sin h}{h}\cdot\frac{1}{\cos x}\cdot\frac{1}{\cos(x+h)}\right\},$$

$$=\frac{1}{\cos^2 x}=\sec^2 x,$$

since $\lim_{h\to 0}\frac{\sin h}{h}=1$, $\lim_{h\to 0}\cos(x+h)=\cos x$,

Hence

$$\boxed{D(\tan x)=\sec^2 x}$$

Remark. We could have also proceeded thus :

Writing $\tan x=\frac{\sin x}{\cos x}$,

and using the formula for the derivative of a quotient of two functions, we have

$$\text{D}(\tan x)=\frac{(\text{D}\sin x)\cos x-(\text{D}\cos x)\sin x}{\cos^2 x},$$

$$=\frac{\cos x \,.\, \cos x-(-\sin x)\,.\,\sin x}{\cos^2 x},$$

$$=\frac{1}{\cos^2 x}=\sec^2 x.$$

3·8·4. Derivative of cot x

If $\quad y=f(x)=\cot x,$

then $\quad f(x+h)=\cot(x+h),$

and
$$\frac{dy}{dx}=\lim_{h\to 0}\frac{\cot(x+h)-\cot x}{h},$$

$$=\lim_{h\to 0}\frac{\cos(x+h)\sin x-\sin(x+h)\cos x}{h\sin(x+h)\sin x},$$

$$=\lim_{h\to 0}\frac{-\sin h}{h\sin(x+h)\sin x},$$

$$=\lim_{h\to 0}\left\{\frac{\sin h}{h}\cdot\frac{-1}{\sin(x+h)\sin x}\right\},$$

$$=-\frac{1}{\sin^2 x}=-\csc^2 x,$$

since $\lim_{h\to 0}\frac{\sin h}{h}=1$, $\lim_{h\to 0}\sin(x+h)=\sin x.$

Hence

$$\boxed{\text{D}(\cot x)=-\csc^2 x}$$

Remarks 1. Since $\cot x=-\tan(x+\pi/2)$, therefore,

$$\text{D}(\cot x)=\text{D}[-\tan(x+\pi/2)],$$

$$=\sec^2(x+\pi/2).\,1,$$

$$=-\csc^2 x.$$

2. Since $\cot x=\dfrac{\cos x}{\sin x}$, applying the rule for finding the derivative of the quotient of two functions, we have

$$\text{D}(\cot x)=\frac{(\text{D}\cos x)\sin x-(\text{D}\sin x)\cos x}{\sin^2 x},$$

$$=\frac{(-\sin x)\sin x-(\cos x)\cos x}{\sin^2 x},$$

$$=-\frac{1}{\sin^2 x}=-\csc^2 x.$$

3. Writing $\cot x$ as $\dfrac{1}{\tan x}$, we have

$$D\cot x=D\left(\frac{1}{\tan x}\right)=-\frac{\sec^2 x}{\tan^2 x}=-\csc^2 x.$$

3·8·5. Derivative of sec x

If $\qquad y=f(x)=\sec x,$

then $\qquad f(x+h)=\sec(x+h),$

and
$$\frac{dy}{dx}=\lim_{h\to 0}\frac{\sec(x+h)-\sec x}{h},$$

$$=\lim_{h\to 0}\frac{\cos x-\cos(x+h)}{h\cos(x+h)\cos x},$$

$$=\lim_{h\to 0}\frac{2\sin(x+\tfrac{1}{2}h)\sin\tfrac{1}{2}h}{h\cos(x+h)\cos x},$$

$$=\lim_{h\to 0}\frac{h\sin\tfrac{1}{2}}{\tfrac{1}{2}h}\cdot\frac{\sin(x+\tfrac{1}{2}h)}{\cos(x+h)\cos x},$$

$$=\frac{\sin x}{\cos^2 x}=\sec x\tan x,$$

since $\displaystyle\lim_{h\to 0}\frac{\sin(\tfrac{1}{2}h)}{\tfrac{1}{2}h}=1$, $\displaystyle\lim_{h\to 0}\sin(x+\tfrac{1}{2}h)=\sin x$, and

$\displaystyle\lim_{h\to 0}\cos(x+h)=\cos x.$

Hence

$$\boxed{D(\sec x)=\sec x\tan x}$$

Remark. Alternatively, we may proceed as follows :

$$D(\sec x)=D\left(\frac{1}{\cos x}\right),$$

$$=\frac{D(1)\cos x-D(\cos x).\,1}{\cos^2 x}$$

$$=\frac{0.\cos x-(-\sin x).\,1}{\cos^2 x}$$

$$=\frac{\sin x}{\cos^2 x}=\sec x\tan x.$$

3·8·6. Derivative of csc x

If $\quad y=f(x)=\csc x,$

then $\quad f(x+h)=\csc(x+h),$

and
$$\frac{dy}{dx}=\lim_{h\to 0}\frac{\csc(x+h)-\csc x}{h},$$
$$=\lim_{h\to 0}\frac{\sin x-\sin(x+h)}{h\sin(x+h)\sin x},$$
$$=\lim_{h\to 0}\frac{-2\cos(x+\frac{1}{2}h)\sin\frac{1}{2}h}{h\sin(x+h)\sin x},$$
$$=\lim_{h\to 0}\left\{\frac{\sin\frac{1}{2}h}{\frac{1}{2}h}\cdot\frac{-\cos(x+\frac{1}{2}h)}{\sin(x+h)\sin x}\right\},$$
$$=\frac{-\cos x}{\sin^2 x}=-\csc x\cot x,$$

since $\lim\limits_{h\to 0}\dfrac{\sin\frac{1}{2}h}{\frac{1}{2}h}=1,\ \lim\limits_{h\to 0}\cos(x+\frac{1}{2}h)=\cos x,$

$\lim\limits_{h\to 0}\sin(x+h)=\sin x.$

Hence

$$\boxed{D(\csc x)=-\csc x\cot x}$$

Remarks. Alternatively, we can proceed as follows :

1. Since $\csc x=-\sec(x+\pi/2)$ therefore,

$$D(\csc x)=D(-\sec(x+\pi/2)),$$
$$=-\sec(x+\pi/2)\tan(x+\pi/2),$$
$$=-(-\csc x)(-\cot x),$$
$$=-\csc x\cot x.$$

2. Since $\csc x=\dfrac{1}{\sin x}$, therefore,

$$D(\csc x)=D\left(\frac{1}{\sin x}\right),$$
$$=\frac{D(1).\sin x-D(\sin x).1}{\sin^2 x},$$
$$=\frac{0.\sin x-\cos x.1}{\sin^2 x},$$
$$=-\frac{\cos x}{\sin^2 x}=-\csc x\cot x.$$

Example 17. *Find the gradient of the tangent to the curve $y=\tan x$ at $x=\pi/4$.*

Solution $\dfrac{dy}{dx}=\dfrac{d}{dx}(\tan x)=\sec^2 x.$

Hence the gradient of the tangent to the curve $y=\tan x$ at $x=\pi/4$ is $\sec^2(\pi/4)=(\sec \pi/4)^2=2.$

Example 18. *Find the rate of change of the function*

$$f: x \to 3 \sin x+4 \cos x \text{ at } x=\pi/2.$$

Solution. $f(x)=3 \sin x+4 \cos x.$

$$f'(x) = \frac{d}{dx}(3 \sin x+4 \cos x),$$

$$=\frac{d}{dx}(3 \sin x)+\frac{d}{dx}(4 \cos x),$$

$$=3 \cos x-4 \sin x.$$

$$\therefore \quad f'(\pi/2)=3 \cos(\pi/2)-4 \sin(\pi/2),$$

$$=-4.$$

Example 19. *Differentiate $y=\sin 10x$ with respect to x.*

Solution. Let $y=\sin u$, where $u=10x$. Now y is differentiable with respect to u and u is differentiable with respect to x. Hence y is differentiable with respect to x, and

$$\frac{dy}{dx}=\frac{dy}{du}\cdot\frac{du}{dx},$$

$$=\frac{d}{du}(\sin u).\ \frac{d}{dx}(10x),$$

$$=(\cos u).\ 10,$$

$$=10 \cos 10x.$$

Example 20. *Differentiate $y=\sec(3x^2+5)$ with respect to x.*

Solution. Here, $y=\sec u$, where $u=3x^2+5$.

$$\therefore \quad \frac{dy}{dx}=\frac{dy}{du}\cdot\frac{du}{dx},$$

$$=\sec u \tan u.\ 6x,$$

$$=6x \sec(3x^2+5) \tan(3x^2+5).$$

Example 21. *Differentiate the function $x \to \tan^2(2x+3)$.*

Solution. $\frac{d}{dx}\{\tan^2(2x+3)\}$

$$=\frac{d}{d(\tan(2x+3))}[\tan^2(2x+3)]\times$$

$$\frac{d}{d(2x+3)}[\tan(2x+3)].\frac{d}{dx}(2x+3),$$

$$=2\tan(2x+3)\sec^2(2x+3).\,2,$$

$$=4\tan(2x+3)\sec^2(2x+3).$$

EXERCISE 3 (*f*)

Differentiate :

1. $\sin(3x+4)$.
2. $\cos^2 x$.
3. $x\tan(4-x^2)$.
4. $\cot^3(2x-1)$.
5. $\sec(2x^2+x+1)$.
6. $x^2\csc(3x-5)$.
7. $\sin^3 x\cos^2 x$.
8. $\sin(\tan x)$.
9. $\sqrt{\cos x}$.
10. $\cos\sqrt{x}$.
11. $\frac{1}{\sin x+\cos x}$.
12. $\tan^2(x^3)$.
13. $x\cos^2(2x-1)$.
14. $(\cot 3x)^{1/3}$.
15. $\csc(x^2)+\csc^2 x$.

3.9. DERIVATIVES OF INVERSE TRIGONOMETRIC FUNCTIONS

Having obtained the derivatives of the trigonometric functions in the preceding section, we shall now use the inverse function theorem to obtain the derivatives of the inverse trigonometric functions.

3.9.1 Derivative of $\sin^{-1}x$

Let $y=f(x)=\sin^{-1}x$,
so that $x=\sin y$.
$=g(y)$, say.

Let us recall that (*i*) f is a strictly increasing function with domain $[-1, 1]$ and range $[-\pi/2, \pi/2]$, (*ii*) g is a strictly increasing function with (suitably restricted) domain $[-\pi/2, \pi/2]$ and range $[-1, 1]$, and that (*iii*) f and g are inverses of each other.

Since $g'(y)=\cos y$ does not vanish anywhere in the open interval $]-\pi/2, \pi/2[$(in fact $g'(y)>0$ throughout this interval), therefore by the inverse function theorem, the function f possesses a derivative for all those values of x for which $-\pi/2<\sin^{-1}x<\pi/2$, *i.e.*, for $-1<x<1$, and

$$f'(x)=\frac{dy}{dx}=\left[\frac{dx}{dy}\right]^{-1}$$
$$=1/[g'(y)]$$

$$=\frac{1}{\cos y}=\frac{1}{\sqrt{1-x^2}}$$

Here $\cos y=\sqrt{1-x^2}$, because $\cos^2 y=1-\sin^2 y=1-x^2$, and $\cos y>0$ whenever $-\pi/2<y<\pi/2$.

Thus

$$\boxed{D(\sin^{-1}x)=\frac{1}{\sqrt{1-x^2}}, \text{ for } -1<x<1}$$

3·9·2. Derivative of cos⁻¹ x

Let $\qquad y=f(x)=\cos^{-1}x,$

so that $\qquad x=\cos y=g(y)$, say.

f is a strictly decreasing function with domain $[-1, 1]$ and range $[0, \pi]$, g is a strictly decreasing function with (suitably restricted) domain $[0, \pi]$ and range $[-1, 1]$, and the functions f and g are inverses of each other.

Since $g'(y)=-\sin y$ does not vanish anywhere in the open interval $]0, \pi[$ (in fact $g'(y)<0$ throughout this interval), therefore by the inverse function theorem, the function f possesses a derivative for all those values of x for which $0<\cos^{-1}x<\pi$, *i.e.*, for $-1<x<1$, and

$$f'(x)=\frac{dx}{dy}=\left[\frac{dx}{dy}\right]^{-1}$$
$$=1/[g'(y)],$$
$$=-\frac{1}{\sin y},$$
$$=-\frac{1}{\sqrt{1-x^2}}.$$

Here $\sin y=\sqrt{1-x^2}$, because $\sin^2 y=1-\cos^2 y=1-x^2$, and $\sin y>0$ whenever $0<y<\pi$.

Thus

$$\boxed{D(\cos^{-1}x)=-\frac{1}{\sqrt{1-x^2}}, \text{ for } -1<x<1.}$$

Remark. We could have also obtained the above formula by differentiating the identity

$$\sin^{-1}x+\cos^{-1}x=\pi/2.$$

3·9·3. Derivative of $\tan^{-1}x$

Let $\quad y=f(x)=\tan^{-1}x,$

so that $\quad x=\tan y=g(y)$, say.

f is a strictly decreasing function with domain **R** and range $]-\pi/2, \pi/2[$, g is strictly increasing in the restricted domain $-]\pi/2, \pi/2[$. The functions f and g are inverses of each other.

Also, $g'(y)=\sec^2 y>0$ for all $y\in\ ]-\pi/2, \pi/2[$.

Therefore by the inverse function theorem, the function f possesses a derivative for all $x\in\mathbf{R}$ (the range of g and domain of f), and

$$f'(x)=\frac{dy}{dx}=\left[\frac{dx}{dy}\right]^{-1}$$
$$=1/[g'(y)].$$
$$=\frac{1}{\sec^2 y},$$
$$=\frac{1}{1+x^2}.$$

Thus

$$\boxed{D(\tan^{-1}x)=\frac{1}{1+x^2},\ \text{for all } x\in\mathbf{R}}$$

3·9·4. Derivative of $\cot^{-1}x$

Differentiating the identity

$$\tan^{-1}x+\cot^{-1}x=\pi/2,$$

we have

$D(\tan^{-1} x)+D(\cot^{-1} x)=D(\pi/2)=0$, for all $x\in\mathbf{R}$,

i.e. $\quad D(\cot^{-1} x)=-D(\tan^{-1} x),$

$$=-\frac{1}{1+x^2},\ \text{for all } x\in\mathbf{R}.$$

Thus

$$\boxed{D(\cot^{-1} x)=\frac{-1}{1+x^2},\ \text{for all } x\in\mathbf{R}}$$

Remark. We should have obtained the above result directly by proceeding in the same way as for obtaining the derivatives of $\sin^{-1} x$, $\cos^{-1} x$ and $\tan^{-1} x$.

3·9·5. Derivative of $\sec^{-1} x$

Since $\sec^{-1} x=\cos^{-1}(1/x)$, when $|x| \geqslant 1$,

therefore by the chain rule for differentiating the composite of two functions, we have

$$D(\sec^{-1} x)=-\frac{1}{\sqrt{1-(1/x)^2}}\frac{d}{dx}(1/x).$$

$$=-\frac{|x|}{\sqrt{x^2-1}}(-1/x^2).$$

$$=\frac{1}{|x|\sqrt{x^2-1}}, \text{ if } |x|>1.$$

Let us note here that while $\sec^{-1} x$ is defined for $|x| \geqslant 1$, $D(\sec^{-1} x)$ does not exist when $x=\pm 1$.

Remark. Note that

$$\sqrt{1-(1/x)^2}>0,$$

and therefore

$$\sqrt{1-(1/x)^2}=\frac{\sqrt{x^2-1}}{|x|}$$

3·9·6. Derivative of $\csc^{-1} x$

Since $\csc^{-1} x=\sin^{-1}(1/x)$, when $|x| \geqslant 1$,

therefore by the chain rule for differentiating the composite of two functions, we have

$$D(\csc^{-1} x)=\frac{1}{\sqrt{1-(1/x)^2}}\frac{d}{dx}(1/x).$$

$$=\frac{-1}{|x|\sqrt{x^2-1}}, |x|>1.$$

Let us note here that while $\csc^{-1} x$ is defined for $|x| \geqslant 1$, $D(\csc^{-1} x)$ does not exist when $x=\pm 1$.

Example 22. *Differentiate*

$$\tan^{-1}\left(\frac{1+x}{1-x}\right).$$

S lution. $D\left(\tan^{-1}\left(\frac{1+x}{1-x}\right)\right)$

$$=\frac{1}{1+\left(\frac{1+x}{1-x}\right)^2}\cdot\frac{d}{dx}\left(\frac{1+x}{1-x}\right),$$

$$=\frac{(1-x)^2}{(1-x)^2+(1+x)^2}\cdot\frac{(1-x)-(1+x)(-1)}{(1-x)^2},$$

$$=\frac{1}{1+x^2}.$$

EXERCISE 3 (g)

Differentiate each of the following :

1. $\sin^{-1}(x/a)$.
2. $\tan^{-1} x^2$.
3. $x \cos^{-1}(2x)$.
4. $\csc^{-1} x^3$.
5. $\cot^{-1}(2x+3)$.
6. $(\sin^{-1} 3x)^3$.
7. $\sqrt{x} \cos^{-1} \sqrt{x}$.
8. $\tan^{-1} \dfrac{2x}{1-x^2}$.
9. $x \csc^{-1}(x/3)$.
10. $[\sec^{-1}(x+1)]^2$.

3·10. USE OF TRANSFORMATIONS IN DIFFERENTIATION

Sometimes a transformation of the function to be differentiated turns out to be quite helpful as a labour saving device. The following examples illustrate the use of transformations in differentiation.

Example 23. *Differentiate*

$$\cos^{-1} \frac{1-x^2}{1+x^2}.$$

Solution. Let $y=f(x)=\cos^{-1}\dfrac{1-x^2}{1+x^2}$.

Substituting $x=\tan\theta$, we have

$$y=\cos^{-1}\frac{1-\tan^2\theta}{1+\tan^2\theta},$$

$$=\cos^{-1}(\cos 2\theta),$$

$$=2\theta=2\tan^{-1} x.$$

$$\therefore \qquad \frac{dy}{dx}=\frac{2}{1+x^2}.$$

Example 24. *Find the derivative of* $\sin^{-1}\dfrac{2x}{1+x^2}$ *with respect to* $\tan^{-1}\dfrac{2x}{1-x^2}$.

Solution. Let $y=\dfrac{2x}{1-x^2}$, ...(1)

and $z=\tan^{-1}\dfrac{2x}{1-x^2}$ ...(2)

Then we have to find $\dfrac{dy}{dz}$.

Substituting $x=\tan\theta$ in (1) and (2), we have

$$y=\sin^{-1}\frac{2\tan\theta}{1+\tan^2\theta},$$

$$=\sin^{-1}(\sin 2\theta),$$

$$=2\theta=2\tan^{-1}x, \qquad \ldots(3)$$

and

$$z=\tan^{-1}\frac{2x}{1-x^2},$$

$$=\tan^{-1}\left(\frac{2\tan\theta}{1-\tan^2\theta}\right),$$

$$=\tan^{-1}(\tan 2\theta),$$

$$=2\theta=2\tan^{-1}x. \qquad \ldots(4)$$

From (3) and (4), we have

$$\frac{dy}{dx}=\frac{2}{1+x^2},$$

$$\frac{dz}{dx}=\frac{2}{1+x^2},$$

so that

$$\frac{dy}{dz}=\frac{dy}{dx}\cdot\frac{dx}{dz},$$

$$=\frac{dy}{dx}\Big/\frac{dz}{dx},$$

$$=1.$$

Remark. Observe that from (3) and (4), we have $y=z$ so that we could have directly got

$$\frac{dy}{dz}=1.$$

Example 25. *Differentiate* $\tan^{-1}[\{\sqrt{(1+x^2)}-1\}/x]$ *with respect to* $\tan^{-1}x$.

Solution. Let $y=\tan^{-1}\dfrac{\sqrt{(1+x^2)}-1}{x}$, $\qquad \ldots(1)$

and

$$z=\tan^{-1}x. \qquad \ldots(2)$$

Then we are required to find $\dfrac{dy}{dz}$.

From (2) we have $x=\tan z$, which when substituted in (1) gives,

$$y=\tan^{-1}\frac{\sqrt{(1+\tan^2 z)}-1}{\tan z},$$

$$=\tan^{-1}\left[\frac{\sec z-1}{\tan z}\right],$$

$$=\tan^{-1}\left[\frac{1-\cos z}{\sin z}\right],$$

$$=\tan^{-1}(\tan z/2),$$

$$=z/2,$$

so that $$\frac{dy}{dz}=\tfrac{1}{2}.$$

EXERCISE 3 (*h*)

Differentiate :

1. $\sin^{-1}\dfrac{2x}{1+x^2}$.
2. $\tan^{-1}\dfrac{2x}{1+x^2}$.
3. $\cos^{-1}(2x^2-1)$.
4. $\cos^{-1}(4x^3-3x)$.
5. $\sin^{-1}(3x-4x^3)$.
6. $\tan^{-1}\dfrac{\sqrt{x}-\sqrt{a}}{1+\sqrt{(ax)}}$.
7. $\tan^{-1}\dfrac{3x-x^3}{1-3x^2}$.
8. $\tan^{-1}\left\{\left(\dfrac{1+\cos x}{1-\cos x}\right)^{1/2}\right\}$.
9. $\sin^{-1}(2ax\sqrt{(1-a^2x^2)})$.
10. $\tan^{-1}\left[\dfrac{x^{1/3}+a^{1/3}}{1-x^{1/3}a^{1/3}}\right]$.
11. $\tan^{-1}\dfrac{\sqrt{(1+x^2)}-1}{x}$.
12. $\tan^{-1}\dfrac{\cos x}{1+\sin x}$.
13. Differentiate $\tan^{-1}\dfrac{2x}{1-x^2}$ with respect to $\cos^{-1}\dfrac{1-x^2}{1+x^2}$.
14. Differentiate $\tan^{-1}\left(\dfrac{a\cos x-b\sin x}{b\cos x+a\sin x}\right)$ with respect to $\tan^{-1}x$.
15. Differentiate $\tan^{-1}\dfrac{\sqrt{(1+x^2)}-\sqrt{(1-x^2)}}{\sqrt{(1+x^2)}+\sqrt{(1-x^2)}}$ with respect to $\cos^{-1}x^2$.

3·11. THE EXPONENTIAL FUNCTION

We shall show that $De^x=e^x$. To establish it we shall need the following result, which we have already proved.

$$\lim_{h\to 0}\frac{e^h-1}{h}=1.$$

We are now ready to prove :

Theorem 3·9. $De^x=e^x$, for all $x\in\mathbf{R}$.

Proof. Let $y=f(x)=e^x$.

Then
$$\begin{aligned}\frac{dy}{dx} &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h},\\ &= \lim_{h\to 0}\frac{e^{x+h}-e^x}{h}\\ &= \lim_{h\to 0}\frac{e^x(e^h-1)}{h},\end{aligned}$$

$$=e^x\left(\lim_{h\to 0}\frac{e^h-1}{h}\right),$$

$$=e^x, \text{ since } \lim_{h\to 0}\frac{e^h-1}{h}=1.$$

The derivative of the composite of the exponential function and a differentiable function h can be obtained by the chain rule, so that

$$De^{h(x)}=e^{h(x)}\,Dh(x)=e^{h(x)}\,h'(x).$$

Example 26. *Differentiate :*

(i) $\dfrac{e^x-e^{-x}}{2}$ (ii) $\dfrac{e^x+e^{-x}}{2}$.

(iii) $\dfrac{e^x-e^{-x}}{e^x+e^{-x}}$. (iv) $\dfrac{e^x+e^{-x}}{e^x-e^{-x}}$.

(v) $\dfrac{2}{e^x+e^{-x}}$. (vi) $\dfrac{2}{e^x-e^{-x}}$.

Solution.

(i) $D\left(\dfrac{e^x-e^{-x}}{2}\right)=\dfrac{e^x+e^{-x}}{2}$.

(ii) $D\left(\dfrac{e^x+e^{-x}}{2}\right)=\dfrac{e^x-e^{-x}}{2}$.

(iii) $$D\left(\frac{e^x-e^{-x}}{e^x+e^{-x}}\right)=\frac{(e^x+e^{-x})\,.\,(e^x+e^{-x})-(e^x-e^{-x})(e^x-e^{-x})}{(e^x+e^{-x})^2},$$

$$=\frac{4}{(e^x+e^{-x})^2}.$$

(iv) $$D\left(\frac{e^x+e^{-x}}{e^x-e^{-x}}\right)=\frac{(e^x-e^{-x})(e^x-e^{-x})-(e^x+e^{-x})(e^x+e^{-x})}{(e^x-e^{-x})^2},$$

$$=\frac{-4}{(e^x+e^{-x})^2}.$$

(v) $$D\left(\frac{2}{e^x+e^{-x}}\right)=2.(-1)\,\frac{1}{(e^x+e^{-x})^2}\cdot(e^x-e^{-x}),$$

$$=\frac{-2(e^x-e^{-x})}{(e^x+e^{-x})^2}.$$

(vi) $$D\left(\frac{2}{e^x-e^{-x}}\right)=2.\frac{-1}{(e^x-e^{-x})^2}\cdot(e^x+e^{-x}),$$

$$=\frac{-2(e^x+e^{-x})}{(e^x-e^{-x})^2}.$$

Example 27. *Differentiate :*

$$e^{\sin x}-2e^{\sqrt{x}}+e^{\cos^{-1} x}.$$

Solution. $D(e^{\sin x}-2e^{\sqrt{x}}+e^{\cos^{-1} x})$.

$$=D(e^{\sin x})-2De^{\sqrt{x}}+De^{\cos^{-1} x},$$

$$=e^{\sin x}\,D(\sin x)-2e^{\sqrt{x}}\,.\,D(\sqrt{x})$$

$$+e^{\cos^{-1} x}D(\cos^{-1} x),$$

$$=e^{\sin x}\cos x-\frac{1}{\sqrt{x}}\;e^{\sqrt{x}}-\frac{1}{\sqrt{(1-x^2)}}\,e^{\cos^{-1} x}.$$

EXERCISE 3 (*i*)

Differentiate :

1. e^{2x+3}.
2. e^{2x^2+5x-7}.
3. $e^{\sin x}$.
4. $e^{\cos^{-1}\sqrt{x}}$.
5. $(e^{2x}+1)^{1/2}$.
6. $\tan(e^{\sqrt{x}})$.
7. $\dfrac{e^x}{\sqrt{(e^{2x}+1)}}$.
8. $x\tan^{-1}(e^{x^2})$.
9. $x^2e^{\sqrt{x}}$.
10. $\cos(\sin^{-1}(\frac{1}{2}e^{\sqrt{x}}))$.

3·12. THE LOGARITHMIC FUNCTION

Let us recall that the natural logarithm function, denoted by *ln*, is defined by

$$y=\ln x \text{ if and only if } x=e^y.$$

The domain of *ln* is $]0, \infty[$ and the range of *ln* is the set **R** of real numbers.

Since the natural logarithm function and the exponential function are inverses of each other, therefore,

$$\ln e^x=x, \text{ for all } x\in \mathbf{R}$$
$$e^{\ln x}=x, \text{ for all } x>0.$$

The following theorem summarizes some of the basic properties of the function *ln*.

Theorem 3·10. *For all positive numbers x and y :*

(*i*) $\ln xy=\ln x+\ln y$,

(*ii*) $\ln\dfrac{x}{y}=\ln x-\ln y$,

(*iii*) $\ln x^r = r \ln x$, for every $r \in \mathbf{R}$,

(*iv*) $\ln x = \ln y$ if and only if $x = y$,

(*v*) $\ln x < \ln y$ if and only if $x < y$.

3·12·1. Derivative of ln x

Let $\qquad y = f(x) = \ln x,$

so that $\qquad x = e^y = g(y)$, say.

Since $g'(y) = e^y \neq 0$ for any $y \in \mathbf{R}$, therefore by the inverse function theorem,

$$\begin{aligned} D(\ln x) = f'(x) &= \frac{dy}{dx}, \\ &= \left(\frac{dx}{dy}\right)^{-1}, \\ &= \frac{1}{e^y}, \\ &= \frac{1}{x}, \text{ for all } x > 0. \end{aligned}$$

Thus

$$\boxed{D(\ln x) = \frac{1}{x},\ x > 0}$$

We can differentiate a composite of the function *ln* and a differentiable function *h* by the chain rule, so that

$$\begin{aligned} D\{\ln h(x)\} &= \frac{1}{h(x)}\ D\ h(x). \\ &= \frac{h'(x)}{h(x)}, \text{ provided } h(x) > 0. \end{aligned}$$

Example 28. *Find* $\quad D \ln (x^2 - 4x + 6)$.

Solution. Since

$$x^2 - 4x + 6 = (x-2)^2 + 2 > 0, \text{ for all } x \in \mathbf{R},$$

therefore $\ln (x^2 - 4x + 6)$ is defined for all $x \in \mathbf{R}$. Moreover,

$$\begin{aligned} D \ln (x^2 - 4x + 6) &= \frac{1}{x^2 - 4x + 6}\, D(x^2 - 4x + 6), \\ &= \frac{2(x-2)}{x^2 - 4x + 6},\ \forall x \in \mathbf{R}. \end{aligned}$$

Example 29. *Find* $\frac{dy}{dx}$, *when* $y = \ln |x^5+1|$.

Solution. $|x^5+1| > 0$ except when $x=-1$.

Also, $|x^5+1| = 0$, when $x=-1$.

Therefore, if $x > -1$, then

$$y = \ln |(x^5+1)|,$$
$$= \ln (x^5+1), \quad \because \ x^5+1 > 0.$$

$$\therefore \quad \frac{dy}{dx} = \frac{5x^4}{x^5+1},$$

If $x < -1$, then $y = \ln \{-(x^5+1)\}$,

$$= \frac{1}{-(x^5+1)} \cdot (-5x^4),$$
$$= \frac{5x^4}{x^5+1}$$

Thus, if $x \neq -1$, then $\frac{dy}{dx} = \frac{5x^4}{x^5+1}$.

Example 30. *Differentiate* :

(*i*) $\ln \{x + \sqrt{(x^2+1)}\}$,

(*ii*) $\ln \{x + \sqrt{(x^2-1)}\}$, $x \geqslant 1$.

(*iii*) $\frac{1}{2} \ln \frac{1+x}{1-x}$, $|x| < 1$.

(*iv*) $\frac{1}{2} \ln \frac{x+1}{x-1}$, $|x| > 1$.

(*v*) $\ln \frac{1+\sqrt{1-x^2}}{x}$, $0 < x < 1$.

(*vi*) $\ln \left[\frac{1}{x} + \frac{\sqrt{1+x^2}}{x}\right]$, $x \neq 0$.

Solution. (*i*) $D \{\ln (x + \sqrt{x^2+1})\}$

$$= \frac{1}{x + \sqrt{(x^2+1)}} \cdot \left\{1 + \tfrac{1}{2} \cdot \frac{2x}{\sqrt{(x^2+1)}}\right\}$$
$$= \frac{1}{\sqrt{(x^2+1)}}.$$

(*ii*) $D \{\ln (x + \sqrt{x^2-1})\}$

$$= \frac{1}{x + \sqrt{(x^2-1)}} \left\{1 + \tfrac{1}{2} \cdot \frac{2x}{\sqrt{(x^2-1)}}\right\},$$
$$= \frac{1}{\sqrt{(x^2-1)}}, \ x > 1.$$

For $x=1$, the derivative does not exist.

(*iii*) Since $|x|<1$, both $x+1$ and $x-1$ are positive. Now

$$\mathrm{D}\left(\tfrac{1}{2}\, ln\frac{1+x}{1-x}\right)=\mathrm{D}\{\tfrac{1}{2}\, ln\,(1+x)-\tfrac{1}{2}\, ln\,(1-x)\},$$

$$=\frac{1}{2(1+x)}-\frac{1}{2(1-x)},$$

$$=\frac{1}{1-x^2},\ |x|<1.$$

(*iv*) Since $|x|>1$, $1/|x|$ is positive. Now,

$$\mathrm{D}\left(\tfrac{1}{2}\, ln\frac{x+1}{x-1}\right)=\mathrm{D}\left(\tfrac{1}{2}\, ln\frac{1+1/x}{1-1/x}\right),$$

$$=\frac{1}{1-1/x^2}\,(-1/x^2),\text{ using (iii) above.}$$

$$=\frac{-1}{x^2-1},\ |x|>1.$$

(*v*) For the given values of x, $1+\sqrt{1-x^2}$ is positive. Therefore,

$$\mathrm{D}\, ln\,\frac{1+\sqrt{1-x^2}}{x}=\mathrm{D}(ln(1+\sqrt{1-x^2})-ln\,x),$$

$$=\frac{\tfrac{1}{2}(-2x)(1-x^2)^{-1/2}}{1+\sqrt{(1-x^2)}}-\frac{1}{x},$$

$$=-\frac{1}{x\sqrt{(1-x^2)}},\ 0<x<1.$$

The derivative does not exist at $x=1$.

(*vi*) If $x>0$, $\mathrm{D}\, ln\left(\frac{1}{x}+\frac{\sqrt{(1+x^2)}}{|x|}\right)$,

$$=\mathrm{D}\ ln\left(\frac{1+\sqrt{(1+x^2)}}{x}\right),$$

$$=\mathrm{D}[ln\,(1+\sqrt{(1+x^2)}-ln\,x],$$

$$=\frac{1}{1+\sqrt{(1+x^2)}}\cdot\tfrac{1}{2}\cdot\frac{2x}{\sqrt{(1+x^2)}}-\frac{1}{x},$$

$$=-\frac{1}{x\sqrt{(1+x^2)}}.$$

If $x<0$, $\mathrm{D}\ ln\left(\frac{1}{x}+\frac{\sqrt{(1+x^2)}}{|x|}\right)$

$$=\mathrm{D}\ ln\left(\frac{\sqrt{1+x^2}-1}{-x}\right),$$

$$=\text{D } ln\,(\sqrt{(1+x^2)}-1)-\text{D } ln\,(-x),$$

$$=\frac{1}{\sqrt{(1+x^2)}-1}\cdot\tfrac{1}{2}\,.\,2x\,.\,(1+x^2)^{-1/2}-\frac{1}{x},$$

$$=\frac{1}{x\sqrt{(1+x^2)}}.$$

The above two results can be written together as

$$\text{D } ln\left(\frac{1}{x}+\frac{\sqrt{(1+x^2)}}{|x|}\right)=-\frac{1}{|x|\sqrt{(1+x^2)}},\ x\neq 0.$$

EXERCISE 3 (*j*)

1. $ln\sqrt{x}$.
2. $ln\,(\cos^2 x)$.
3. $ln\left|\dfrac{1+x^2}{1-x^2}\right|$.
4. $ln\sqrt{(x^2+4)}$.
5. $ln\sqrt{e^x+1}$.
6. $e^{x^2}\,ln\,x$.
7. $[ln\,(x^2+1)]^2$.
8. $\dfrac{1}{ln\,x}$.

3·13. DERIVATIVE OF THE EXPONENTIAL FUNCTION WITH BASE a (>0)

We are now in a position to obtain the derivative of the function defined by

$$y=a^x,\ x\in\mathbf{R},$$

a being any fixed positive real number. In fact,

$$y=a^x \Leftrightarrow \ln y=x\ln a,$$

$$\Leftrightarrow y=e^{x\,ln\,a}$$

Therefore, $\dfrac{dy}{ex}=\text{D}\,(e^{x\,ln\,a})$

$$=e^{x\,ln\,a}\,(\ln a),$$

$$=a^x\,(\ln a).$$

Thus

$$\boxed{\text{D}a^x=a^x\ln a,\ a>0}$$

3·14. LOGARITHMS TO BASES OTHER THAN e

You have already read about logarithms to base 10, and logarithms to base e (natural logarithms). We can, in fact, talk of logarithms to any base $a>0$ but different from 1, very much in the same way as we can talk of the exponential function to any base $a>0$ (but other than 1, the case $a=2$ being trivial).

If a and x are positive real numbers with $a \neq 1$, then

$$x = a^y \Leftrightarrow \ln x = y \ln a,$$

$$\Leftrightarrow y = \frac{\ln x}{\ln a},$$

so that for each real number $x > 0$, there exists a unique real number y such that

$$x = a^y.$$

Definition 3·2. *Let a be a positive real number, $a \neq 1$. The logarithm function to the base a, denoted by log a, is defined by setting*

$$y = \log_a x \text{ if and only if } x = a^y.$$

In view of the above definition, we find that :

(*i*) the function $\log_e$ is simply the function ln.

(*ii*) $\log_a x = \dfrac{\ln x}{\ln a}$.

(*iii*) $\log_a a^y = y$, for all $y \in \mathbf{R}$.

$a^{\log_a x} = x$, for all $x > 0$.

(*iv*) $\log_a e = \dfrac{1}{\ln a}$.

(*v*) $\log_a x = (\log_a e) \ln x$.

The last of the above relations serves to obtain the logarithm of any positive number x to base a in terms of its logarithm to base e.

To differentiate the logarithmic function to the base a, we have from (*v*),

$$\mathrm{D}(\log_a x) = \log_a e \, \mathrm{D}(\ln x),$$

$$= \log_a e \cdot \frac{1}{x}.$$

Thus

$$\boxed{\mathrm{D}(\log_a x) = \log_a e . \frac{1}{x}}$$

We can differentiate a composite of the function 'logarithm to base a, and a differentiable function h by the chain rule, so that

$$\mathrm{D} \log_a h(x) = \log_a e . \frac{1}{h(x)} . \mathrm{D}h(x)$$

$$= \frac{h'(x) \log_a e}{h(x)}.$$

3·15. THE POWER FUNCTION

If r is an irrational number, and x is any positive real number, then by definition x^r is a unique real number. This enables us to define the power function for irrational exponents as well.

Definition 3·3. *If r is a real number, then the function $x \to x^r$ for all $x>0$, is called the power function.*

We have already seen that if r is a rational number and $x \neq 0$, then

$$Dx^r = rx^{r-1}.$$

We shall show that this formula holds even when r is irrational.

Let r be any irrational number, $x>0$, and let

$$y = x^r.$$

Writing $\quad x^r = e^{r \ln x},$

we have
$$\frac{dy}{dx} = D(e^{r \ln x})$$
$$= e^{r \ln x}\, D(r \ln x),$$
$$= e^{r \ln x}\left(\frac{r}{x}\right),$$
$$= rx^{r-1}.$$

Thus
$$\boxed{D^r = rx^{r-1}}$$

The above formula holds for all $x \in \mathbf{R}$ if r is a non-zero integer, and for all $x>0$ if r be any non-zero real number. (Of course, if $r=0$ then $Dx^r=0$ for all $x \in \mathbf{R}$).

Example 31. *If $f(x) = 5^{x^2}$, find $f'(x)$.*

Solution. $\dfrac{d}{dx}(5^{x^2}) = (5^{x^2} \ln 5)\dfrac{d}{dx}(x^2),$

$$= 2x.5^{x^2} \ln 5.$$

Example 32. *Find $D \log_2 (3x^2-5x+8)$.*

Solution. Since $\log_2 (3x^2-5x-8)$

$$= (\log_2 e) \ln (3x^2-5x+8),$$

therefore, $D \log_2 (3x^2-5x+8)$

$$= (\log_2 e)\, D[\ln(3x^2-5x+8)],$$
$$= \frac{(\log_2 e)(6x-5)}{3x^2-5x+8}.$$

Example 33. *If* $y=\sin (x^2+1)^{\sqrt{2}}$,

find $\dfrac{dy}{dx}$.

Solution. $\dfrac{dy}{dx}=\cos\ (x^2+1)^{\sqrt{2}}\ d\left[(x^2+1)^{\sqrt{2}}\right]$

$=\left[\cos(x^2+1)^{\sqrt{2}}\right]\left[\sqrt{2}(x^2+1)^{\sqrt{2}-1}\ (2x)\right]$,

$=2\sqrt{2}x(x^2+1)^{\sqrt{2}-1}\cos(x^2+1)^{\sqrt{2}}$.

EXERCISE 3 (*k*)

Differentiate each of the following :

1. 10^{x^2}.
2. $5^{\sin x}$.
2. $2^{\cos x^2}$,
4. $\log_5 (2x+1)$.
5. $x \log_5 \tan^2 x$.
6. $\sin x \log_{10} (x^2+1)$.
7. $(x^2+4)^{\sqrt{3}}$.
8. $(\sqrt{x}+2)^{e+1}$.

3·16. LOGARITHMIC DIFFERENTIATION

When we have to find the derivative of a function of the type

$$y=\frac{u_1u_2\cdots\cdots u_k}{v_1v_2\cdots\cdots v_l},$$

or of the type $y=u^v$, where $u_1, u_2,\cdots\cdots, u_k, v_1, v_2\cdots\cdots, v_l, u, v$ are function of x, then it is convenient to take logarithms before differentiating. This process is called *logarithmic differentiation.*

Type I.

$$\text{Let } y=f\ (x)=\frac{u_1u_2\ldots\ldots u_k}{v_1v_2\cdots\cdots v_l},$$

where $u_1, u_2,\ldots\ldots, u_k, v_1, v_2,\cdots\cdots v_l$ are functions of x. Taking logarithms, we have

$$\ln f(x)=(\ln u_1+\ln u_2+\ldots\ldots+\ln u_k)-(\ln v_1+v_2+\cdots\cdots+\ln v_l).\ \cdots(1)$$

$$\text{Since } D(\ln f(x))=\frac{1}{f(x)}\ D\ (f(x))=\frac{f'(x)}{f(x)},$$

$$D(\ln u_1)=\frac{u'_1}{u_1},\ldots\ldots,$$

therefore, differentiating both sides of (1), we have

$$\frac{f'(x)}{f(x)}=\left(\frac{u'_1}{u_1}+\frac{u'_2}{u_2}+\ldots\ldots+\frac{u'_k}{u_k}\right)$$

$$-\left(\frac{v'_1}{v_1}+\frac{v'_2}{v_2}+\dots\dots+\frac{v'_l}{v_l}\right)$$

or
$$\frac{dy}{dx}=y\left\{\frac{u'_1}{u_1}+\frac{u'_2}{u_2}\right.$$
$$\left.+\dots+\frac{u'_k}{u_k}-\frac{v'_1}{v_1}-\frac{v'_2}{v_2}-\dots\dots-\frac{v'_1}{v_2}\right\},$$

Example 34. *Find the derivative of*

$$(x-1)^2\,(x-2)^3\,(x-3)^6.$$

Solution. Writing
$$y=(x-1)^2\,(x-2)^3\,(x-3)^6,$$
we have
$$\ln y=2\ln(x-1)+3\ln(x-2)+6\ln(x-3).$$

Differentiating throughout with respect to x, we have
$$\frac{1}{y}\,\frac{dy}{dx}=\frac{2}{x-1}+\frac{3}{x-2}+\frac{6}{x-3},$$
so that
$$\frac{dy}{dx}=y\left\{\frac{2}{x-1}+\frac{3}{x-2}+\frac{6}{x-3}\right\}.$$

Example 35. *Find the derivative of*

$$\frac{(x+1)^{1/2}\,(2x-3)^{5/6}}{(3x+2)^{3/4}\,(5x-4)^{2/3}}.$$

Solution. Writing
$$y=\frac{(x+1)^{1/2}\,(2x-3)^{5/6}}{(3x+2)^{3/4}\,(5x-4)^{2/3}},$$
We have
$$\ln\ y=\frac{1}{2}\ln(x+1)+\frac{5}{6}\ln(2x-3)-\frac{3}{4}\ln(3x+2)$$
$$-\frac{2}{3}\ln(5x-4) \qquad \text{...(I)}$$

Differentiating (I) throughout with respect to x, we have
$$\frac{1}{y}\,\frac{dy}{dx}=\frac{1}{2(x+1)}+\frac{5}{3(2x-3)}-\frac{9}{4(3x+2)}-\frac{10}{3(5x-4)},$$
so that
$$\frac{dy}{dx}=y\left\{\frac{1}{2(x+1)}+\frac{5}{3(2x-3)}+\frac{9}{4(3x+2)}-\frac{10}{3(5x-4)}\right\}.$$

Type II.

Let $\qquad y=f\,(x)=[g(x)]^{h(x)},$
where g, h are both differentiable functions of x.

Taking logarithms of both sides, we have

$$\ln f(x)=h(x)\ln[g(x)]. \quad ...(1)$$

Differentiating both sides of (1), and applying the rule for differentiation of the product of two functions, we have

$$D \ln f(x)=[Dh(x)] \ln g(x)+h(x)D[\ln g(x)] \quad ...(2)$$

Since $D \ln f(x)=\frac{f'(x)}{f(x)}$,

therefore $\frac{f'(x)}{f(x)}=h'(x) \ln g(x)+h(x)\frac{g'(x)}{g(x)}$,

i.e., $$f'(x)=f(x)\left[h'(x) \ln g(x)+h(x)\frac{g'(x)}{g(x)}\right].$$

Thus, $\frac{dy}{dx}=y\left[h'(x) \ln g(x)+h(x)\frac{g'(x)}{g(x)}\right]$.

Example 36. *If* $y=x^{\sin x}$, find $\frac{dy}{dx}$.

Solution. Taking logarithms of both sides of the relation

$$y=x^{\sin x},$$

we have

$$\ln y=\sin x \ln x. \quad ...(1)$$

Differentiating (1) throughout with respect to x, we have

$$\frac{1}{y}\frac{dy}{dx}=\cos x \ln x+\frac{\sin x}{x},$$

or $$\frac{dy}{dx}=y\left\{\cos x \ln x+\frac{\sin x}{x}\right\}.$$

Example 37. *Find the derivative of*

$$(\sin x)^{\tan x}+(\tan x)^{\sin x}.$$

Solution. Let

$$y=(\sin x)^{\tan x}+(\tan x)^{\sin x}, \quad ...(1)$$

$$g(x)=(\sin x)^{\tan x}, \quad ...(2)$$

$$h(x)=(\tan x)^{\sin x}, \quad ...(3)$$

so that $$y=g(x)+h(x),$$

whence $$\frac{dy}{dx}=g'(x)+h'(x). \quad ...(4)$$

We shall obtain $g'(x)$ and $h'(x)$ by logarithmic differentiation, and then substitute in (4) to get $\frac{dy}{dx}$.

Taking logarithms of both sides of (2), we have

$$\ln g(x)=\tan x \ln \sin x.$$

By differentiating both sides with respect to x, we have

$$\frac{g'(x)}{g(x)}=\sec^2 x \ln \sin x+1,$$

so that $$g'(x)=(\sin x)^{\tan x}[\sec^2 x \ln \sin x+1] \qquad ...(5)$$

Again, taking logarithms of both sides of (3), we have

$$\ln h(x)=\sin x \ln \tan x.$$

Differentiating both sides with respect to x, we have

$$\frac{h'(x)}{h(x)}=\cos x \ln \tan x+\sin x \,\frac{\sec^2 x}{\tan x},$$

$$=\cos x \ln \tan x+\sec x,$$

so that

$$h'(x)=(\tan x)^{\sin x}[\cos x \ln \tan x+\sec x]. \qquad ...(6)$$

From (4), (5) and (6), we have

$$\frac{dy}{dx}=(\sin x)^{\tan x}[\sec^2 x \ln \sin x+1]$$
$$+(\tan x)^{\sin x}[\cos x \ln \tan x+\sec x].$$

EXERCISE 3 (*l*)

Differentiate :

1. $(x-1)^2(x-2)\sin x$.
2. $\sin x \cos^{-1} x \ln x$.
3. $\dfrac{x^{\frac{1}{2}}(1-2x)^{\frac{1}{3}}}{(x+2)^{\frac{1}{4}}}$.
4. $\dfrac{1}{(x+1)(x+2)(x+3)}$.
5. $x^{(x^x)}$.
6. $(x^x)^x$.
7. $x^{\sin^{-1}x}$.
8. $(\sin x)^{\cos^{-1} x}$.
9. $(\tan x)^x+x^{\tan x}$.
10. $(\tan x)^{\cot x}+(\cot x)^{\tan x}$.
11. $(\sin x)^{\cos x}+(\cos x)^{\sin x}$.
12. $x^x+(\tan x)^{\ln x}$.

3·17. DIFFERENTIATION OF IMPLICITY DEFINED FUNCTIONS

Most of the functions that we have discussed so far have been explicitly defined by an algebraic equation. For example, the equation

$$y=x^2+1,$$

defines a function f, where $f(x)=x^2+1$.The graph of the function f is simply the graph of the given equation.

All functions are not, however, defined in an explicit manner. For example, the equation

$$y^4-x^4+xy=0,$$

cannot be easily solved for y in terms of x (or, for x in terms of y). However, there might exist a function f such that the equation

$$[f(x)]^4-x^4+xf(x)=0,$$

is true for every x in the domain of f. Such a function is said to be defined implicitly by the given equation.

The derivative of a function defined implicitly by an equation in x and y can often be found without explicitly solving the equation for y in terms of x. The process of finding the derivative in such a case is called **implicit differentiation.**

Example 38. *Find* $\frac{dy}{dx}$ *if*

$$x^3+y^3-6xy=0.$$

Solution. Differentiating the relation

$$x^3+y^3-6xy=0,$$

throughout with respect to x, we have

$$3x^2+3y^2\frac{dy}{dx}-6\left(y+x\frac{dy}{dx}\right)=0.$$

or
$$(3y^2-6x)\frac{dy}{dx}=6y-3x^2,$$

whence
$$\frac{dy}{dx}=\frac{2y-x^2}{y^2-2x}.$$

Remark. Here y^3 is a function of y, and y itself is a function of x. Therefore y^3 has been differentiated by the chain rule. Also, $6xy$ has been differentiated by the rule for the derivative of the product of two functions.

Example 39. *Find* $\frac{dy}{dx}$, *if x and y are related by the equation*

$$x^2 \sin y=y^2 \sin x.$$

Solution. Differentiating both sides with respect to x,

$$\frac{d}{dx}(x^2 \sin y)=\frac{d}{dx}(y^2 \sin x),$$

or $$\frac{d}{dx}(x^2).\sin y+x^2\frac{d}{dx}(\sin y)=\frac{d}{dx}(y).\sin x+y^2\frac{d}{dx}(\sin x),$$

or $$2x \sin y+x^2 \cos y\frac{dy}{dx}=2y\frac{dy}{dx}\sin x+y^2 \cos x,$$

or $(x^2 \cos y-2y \sin x)\dfrac{dy}{dx}=y^2 \cos x-2x \sin y,$

or
$$\frac{dy}{dx}=\frac{y^2 \cos x-2x \sin y}{x^2 \cos y-2y \sin x}.$$

Example 40. *Two differentiable functions are defined by the equation*

$$x^2+y^2=1,$$

of a circle, namely, those defined by the equations,

$$y=\sqrt{(1-x^2)} \text{ and } y=-\sqrt{(1-x^2)}.$$

Find the derivative of each function.

Solution. If f denotes either of the two functions, then

$$x^2+f^2(x)=1,$$

for every x in $[-1, 1]$, the domain of f. By implicit differentiation,

$$D(x^2+f^2(x))=D(1),$$

or
$$2x+2f(x)\, Df(x)=0.$$

or
$$Df(x)=-\frac{x}{f(x)}$$

whenever $\quad f(x)\neq 0.$

Since $f(-1)=f(1)=0$, therefore $-1, 1$ must be excluded from the domain of Df.

By the above formula for Df, we have

$$D\sqrt{(1-x^2)}=-\frac{x}{\sqrt{(1-x^2)}}, \text{ and}$$

$$D(-\sqrt{(1-x^2)})=\frac{x}{\sqrt{(1-x^2)}}.$$

Remark. The above example illustrates the fact that implicit differentiation gives the derivative of every differentiable function defined by the given equation.

EXERCISE 3 (*m*)

Find $\dfrac{dy}{dx}$ for each of the implicitly defined functions below :

1. $ax^2+2hxy+by^2=1.$
2. $ax^2+2hxy+by^2+2gx+2fy+c=0.$
3. $y=x^y.$
4. $x=y \ln (xy).$
5. $y=x \ln \dfrac{y}{a+bx}.$

6. $x^m y^n=(x+y)^{m+n}$.

7. $y \cos x=x \cos y$.

8. $x^5 \sin y+y^5 \sin x=0$.

9. $(\cos x)^y=(\sin y)^x$.

10. $x^y+y^x=c$.

11. If $x^y=e^{x-y}$, prove that

$$\frac{dy}{dx}=\frac{\ln x}{(1+\ln x)^2}.$$

12. If $\sin y=x \sin (a+y)$, prove that

$$\frac{dy}{dx}=\frac{\sin^2 (a+y)}{\sin a}.$$

3·18. DIFFERENTIATION OF FUNCTIONS DEFINED IN TERMS OF A PARAMETER

Sometimes x and y are both expressed in terms of a third variable, say t, which is usually called a parameter. In such cases $\frac{dy}{dx}$ can be directly obtained without first eliminating t and finding the relation connecting x and y. Thus, if

$$x=f(t) \text{ and } y=g(t),$$

then

$$\frac{dy}{dx}=\frac{dy}{dt}\cdot\frac{dt}{dx},$$

$$=\frac{dy}{dt}\Big/\frac{dx}{dt}$$

$$=\frac{g'(t)}{f'(t)}$$

provided $f'(t)\neq 0$.

Example 41. *If $x=a(t-\sin t)$, $y=a(1-\cos t)$ find $\frac{dy}{dx}$.*

Solution. Here

$$\frac{dx}{dt}=a(1-\cos t),$$

$$\frac{dy}{dt}=a \sin t,$$

so that

$$\frac{dy}{dx}=\frac{\frac{dy}{dt}}{\frac{dx}{dt}}=\frac{a \sin t}{a(1-\cos t)}=\cot\left(\frac{t}{2}\right),$$

provided t is not an integral multiple of 2π.

EXERCISE 3 (*n*)

Find $\frac{dy}{dx}$ in each of the following cases :

1. $x=a\cos\theta,\ y=a\sin\theta.$
2. $x=a\cos\theta,\ y=b\sin\theta.$
3. $x=a\sec\theta,\ y=b\tan\theta.$
4. $x=a\cos^3\theta,\ y=b\sin^3\theta.$
5. $x=3\cos\theta-\cos 3\theta,\ y=3\sin\theta-\sin 3\theta.$
6. $x=a\left(\cos\theta+ln\tan\frac{\theta}{2}\right),\ y=a\sin\theta.$
7. $x=a(\theta+\sin\theta),\ y=a(1-\cos\theta).$
8. $x=at^2,\ y=2at.$
9. $x=\sin t\sqrt{(\cos 2t)},\ y=\cos t\sqrt{(\cos 2t)}.$
10. $x=\frac{\sin^3 t}{\sqrt{(\cos 2t)}},\ y=\frac{\cos^3 t}{\sqrt{(\cos 2t)}},$ at $t=\frac{\pi}{6}.$

3·19. SUCCESSIVE DIFFERENTIATION

If f' is the derivative of a derivable function f, then f' is called the *first derivative* of f. If f' is a derivable function, then its derivative is denoted by f'' and is called the *second derivative* of f. Similarly, the third derivative of f is denoted by f''', and so on.

In terms of the D notation for the derivative, the first, second, third derivatives of f etc. are denoted by $Df,\ D^2f,\ D^3f,...$

If $y=f(x)$, then $\frac{dy}{dx},\ \frac{d^2y}{dx^2}$ denote the first derivative, the second derivative,.........of y, respectively. Sometimes we use the symbols $y_1,\ y_2,$ .... to denote the successive derivatives of y.

We shall discuss various methods of obtaining the nth derivative of a function where n is an arbitrary positive integer. First, we shall illustrate the technique of obtaining second derivatives by means of some examples.

Example 42. *Find the second derivative of* $\frac{ax+b}{cx+d}$.

Solution. Writing

$$y=\frac{ax+b}{cx+d},\ \text{we have}$$

$$\frac{dy}{dx}=\frac{a(cx+d)-c(ax+b)}{(cx+d)^2},$$

$$=\frac{ad-bc}{(cx+d)^2},$$

$$\frac{d^2y}{dx^2}=(ad-bc)\,.\,(-2)(cx+d)^{-3}\,.\,c,$$

$$=\frac{-2c(ad-bc)}{(cx+d)^2}.$$

Example 43. *Find the second derivative of sin 3x cos 5x.*

Solution. Writing

$$y=\sin 3x \cos 5x=\tfrac{1}{2}(\sin 8x-\sin 2x),$$

we have $y_1=\frac{1}{2}(8\cos 8x-2\cos 2x)$,

$$y_2=\tfrac{1}{2}(-64\sin 8x+4\sin 2x),$$

$$=-32\sin 8x+2\sin 2x.$$

Example 44. $y=e^{ax}\sin(bx+c)$, *show that*

$$\frac{d^2y}{dx^2}=r^2e^{ax}\sin(bx+c+2\phi),$$

where $r=\sqrt{(a^2+b^2)},\ \phi=\tan^{-1}(b/a)$.

Solution. *If* $y=e^{ax}\sin(bx+c)$,

then $\dfrac{dy}{dx}=ae^{ax}\sin(bx+c)+be^{ax}\cos(bx+c)$. ...(1)

Let $a=r\cos\phi,\ b=r\sin\phi,$ ...(2)

so that $r^2=a^2+b^2,\ \tan\phi=b/a,$

i.e., $r=\sqrt{(a^2+b^2)},\ \phi=\tan^{-1}(b/a)$. ...(3)

Substituting the values of a and b from (2) in (1), we have

$$\frac{dy}{dx}=re^{ax}\{\sin(bx+c)\cos\phi+\cos(bx+c)\sin\phi\},$$

$$=re^{ax}\sin(bx+c+\phi). \qquad ...(4)$$

Differentiating (4) throughout with respect to x, and using the same argument as above, we have

$$\frac{d^2y}{dx^2}=r^2e^{ax}\sin(bx+c+2\phi),$$

where r and ϕ are given by (3).

Example 45. *If* $y=a\cos(\ln x)+b\sin(\ln x)$, *show that*

$$x^2\frac{d^2y}{dx^2}+x\frac{dy}{dx}+y=0.$$

Solution. Since,

$$y=a\cos(\ln x)+b\sin(\ln x),$$

therefore, $\dfrac{dy}{dx}=\left[-a\sin(\ln x)\cdot\dfrac{1}{x}+b\cos(\ln x)\right]\cdot\dfrac{1}{x}$,

so that $x\dfrac{dy}{dx}=-a\sin(\ln x)+b\cos(\ln x)$. ...(1)

Differentiating (1) throughout with respect to x, we have

$$x\frac{d^2y}{dx^2}+\frac{dy}{dx}=\left[-a\cos(\ln x)\cdot\frac{1}{x}+\Big((-b\sin(\ln x)\Big)\right]\cdot\frac{1}{x},$$

$$=-\frac{1}{x}\Big\{a\cos(\ln x)+b\sin(\ln x)\Big\},$$

$$=-\frac{y}{x},$$

so that

$$x^2\frac{d^2y}{dx^2}+x\frac{dy}{dx}+y=0.$$

Example 46. *If $x=a(\theta-\sin\theta)$, $y=a(1-\cos\theta)$, find $\dfrac{d^2y}{dx^2}$.*

Solution. Since

$$x=a(\theta-\sin\theta),$$
$$y=a(1-\cos\theta),$$

therefore

$$\frac{dx}{d\theta}=a(1-\cos\theta),$$

$$\frac{dy}{d\theta}=a\sin\theta,$$

and

$$\frac{dy}{dx}=\frac{dy}{d\theta}\Big/\frac{dx}{d\theta}=\frac{\sin\theta}{1-\cos\theta}.$$

Therefore,

$$\frac{d^2y}{dx^2}=\frac{d}{d\theta}\left(\frac{dy}{dx}\right)\frac{d\theta}{dx},$$

$$=\frac{d}{d\theta}\left(\frac{\sin\theta}{1-\cos\theta}\right)\Big/\frac{dx}{d\theta},$$

$$=\frac{\cos\theta(1-\cos\theta)-(\sin\theta)\sin\theta}{(1-\cos\theta)^2}\cdot\frac{1}{a(1-\cos\theta)},$$

$$=\frac{-1}{a(1-\cos\theta)^2}.$$

Example 47. *If $ax^2+2hxy+by^2=1$, prove that*

$$\frac{d^2y}{dx^2}=\frac{h^2-ab}{(hx+by)^3}.$$

Solution. Differentiating the relation

$$ax^2+2hxy+by^2=1$$

throughout with respect to x, we have

$$2ax+2h\left(y+x\frac{dy}{dx}\right)+2by\frac{dy}{dx}=0,$$

or $$(ax+hy)+(hx+by)\frac{dy}{dx}=0,$$

so that $$\frac{dy}{dx}=-\frac{ax+hy}{hx+by}. \qquad ...(1)$$

Differentiating both sides of (1) with respect to x, we have

$$\frac{d^2y}{dx^2}=-\frac{\left(a+h\frac{dy}{dx}\right)(hx+by)-\left(h+b\frac{dy}{dx}\right)(ax+hy)}{(hx+by)^2},$$

$$=\frac{(h^2-ab)\left(y-x\frac{dy}{dx}\right)}{(hx+by)^2},$$

$$=\frac{h^2-ab}{(hx+by)^2}\left\{y+\frac{x(ax+hy)}{hx+by}\right\}$$

$$=\frac{h^2-ab}{(hx+by)^3}\cdot(ax^2+2hxy+by^2),$$

$$=\frac{h^2-ab}{(hx+by)^3}, \text{ since } ax^2+2hxy+by^2=1.$$

EXERCISE 3 (*o*)

Find the second derivatives of :

1. $(2x+3)^4$.
2. e^{4x}.
3. $\frac{1}{5x+4}$.
4. $\ln(4-3x)$.
5. $\ln\{(ax+b)/(cx+d)\}$.
6. $\sin^{-1} x$.
7. $\sin 2x \sin 4x$.
8. $\tan x+\cot x$.
9. If $y=e^{ax}\sin bx$, find $\frac{d^2y}{dx^2}$.
10. If $y=e^{ax}\cos(bx+c)$, show that $y_2=r^2 e^{ax}\cos(bx+c+2\phi)$ where $r=\sqrt{(a^2+b^2)}$, $\phi=\tan^{-1}(b/a)$.
11. If $y=\sin(\sin x)$, prove that

$$\frac{d^2y}{dx^2}+\tan x\frac{dy}{dx}+y\cos^2 x=0.$$

12. If $y=Ae^{px}+Be^{qx}$, show that

$$\frac{d^2y}{dx^2}-(p+q)\frac{dy}{dx}+pqy=0.$$

13. If $y=(a+bt)e^{nt}$, show that

$$\frac{d^2y}{dt^2}-2n\frac{dy}{dt}+n^2y=0.$$

14. If $y=A\cos nx+b\sin nx$, prove that

$$\frac{d^2y}{dx^2}+n^2y+0.$$

15. If $p^2=a^2\cos^2\theta+b^2\sin^2\theta$, show that

$$\frac{d^2p}{d\theta^2}+p=\frac{a^2b^2}{p^3}.$$

16. If $ax^2+2hxy+by^2+2gx+2fy+c=0$, prove that

$$\frac{d^2y}{dx^2}=\frac{abc+2fgh-af^2-bg^2-ch^2}{(hx+by+f)^3}.$$

17. If $x^3+y^3-3axy=0$, show that

$$\frac{d^2y}{dx^2}=-\frac{2a^3xy}{(y^2-ax)^3}.$$

18. If $\sin(x+y)=py$, where p is a constant, prove that

$$\frac{d^2y}{dx^2}=-y\left(1+\frac{dy}{dx}\right)^3.$$

19. If $x=a\cos\theta$, $y=b\sin\theta$, find $\frac{d^2y}{dx^2}$.

20. If $x=a(\cos\theta+\theta\sin\theta)$ $y=a(\sin\theta-\theta\cos\theta)$, find $\frac{d^2y}{dx^2}$.

21. If $x=2\cos t-\cos 2t$, $y=2\sin t-\sin 2t$, find the value of $\frac{d^2y}{dx^2}$ when $t=\frac{\pi}{2}$.

3·20. nth DERIVATIVES OF SOME STANDARD FUNCTIONS

We shall show that it is possible to determine a formula for the nth derivative of a function in some simple cases, where n is an arbitrary positive integer.

3·20·1. To find the nth derivative of $(ax+b)^m$.

Let $y=(ax+b)^m$

Then $y_1=ma(ax+b)^{m-1}$

$y_2=m(m-1)\,a^2\,(ax+b)^{m-2}$

$$y_3=m(m-1)(m-2)\,a^3\,(ax+b)^{m-3}$$

$$\begin{array}{ccc} \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{array}$$

In general, $y_n=m(m-1)\cdots(m-n+1)\,a^n\,(ax+b)^{m-n}$

Remarks 1. In case m is a positive integer, the above formula can be written in the form

$$y_n=\frac{m\,!}{(m-n)\,!}\,a^n\,(ax+b)^{m-n}.$$

2. The above formula holds for all positive integral values of n if m is not a positive integer. If m is a positive integer then the formula holds for all $n \leqslant m$. For $n=m$, we have $y_m=(m\,!)\,a^m$. Also, $y_n=0$ for all $n>m$.

Corollaries 1. If $y=x^m$, then

$$y_n=m(m-1)\ldots\ldots(m-n+1)\,x^{m-n}.$$

2. If $y=\dfrac{1}{ax+b}$,

$$y_n=\frac{(-1)^n\,(n\,!)\,a^n}{(ax+b)^{n+1}}.$$

3·20·2. To find the nth derivative of ln (ax+b).

Let $y=\ln(ax+b)$

Then $y_1=\dfrac{a}{ax+b}=a(ax+b)^{-1}$

$$y_2=(-1)\,a^2\,(ax+b)^{-2}$$

$$y_3=(-1)^2\,2a^3\,(ax+b)^{-3}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

In general, $y_n=(-1)^{n-1}\,\dfrac{a^n(n-1)\,!}{(ax+b)^n}$.

Corollary. If $y=\ln x$, then

$$y_n=(-1)^{n-1}\frac{(n-1)\,!}{x^n}.$$

3·20·3. To find the nth derivative of a^{mx}.

Let $y=a^{mx}$

Then, $y_1=ma^{mx}\ln a$

$$y_2=m^2\,a^{mx}\,(\ln a)^2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

In general, $y_n=m^n\,a^{mx}\,(\ln a)^n$.

Corolleries 1. If $y=a^x$, then by putting $m=1$ in the above formula, we get

$$y_n=a^x\,(\ln a)^n,$$

2. If $y=e^x$, then by putting $m=1$ and $a=e$ in the above formula, we get

$$y_n=e^x.$$

3. If $y=e^{nx}$, then

$$y_n=m^n\,e^{mx}.$$

3·20·4. To find the nth derivatives of sin (ax+b) and cos (ax+b).

First, let $y=\sin(ax+b)$.

Then, $y_1=a\cos(ax+b)=a\sin\left(ax+b+\dfrac{\pi}{2}\right)$

$$y_2=a^2\cos\left(ax+b+\frac{\pi}{2}\right)=a^2\sin(ax+b+\pi)$$

$$y_3=a^3\cos(ax+b+\pi)=a^3\sin\left(ax+b+\frac{3\pi}{2}\right)$$

....

In general, $y_n=a^n\sin\left(ax+b+\dfrac{n\pi}{2}\right)$.

Similarly, if $y=\cos(ax+b)$, then

$$y_n=a^n\cos\left(ax+b+\frac{n\pi}{2}\right).$$

Corollary. If $y=\sin x$, then $y_n=\sin\left(x+\dfrac{n\pi}{2}\right)$

and if $y=\cos x$, then $y_n=\cos\left(x+\dfrac{n\pi}{2}\right)$.

3·20·5. To find the nth derivatives of e^{ax} sin (bx+c) and e^{ax} cos (bx+c).

First, let $y=e^{ax}\sin(bx+c)$.

Then, $y_1=ae^{ax}\sin(bx+c)+be^{ax}\cos(bx+c)$.

Put $a=r\cos\theta$ and $b=r\sin\theta$.

Then, $r^2=a^2+b^2$ and $\tan\theta=\dfrac{b}{a}$.

With these values,

$$y_1=re^{ax}[\sin(bx+c)\cos\theta+\cos(bx+c)\sin\theta]$$
$$=re^{ax}\sin(bx+c+\theta).$$

Similarly, $y_2=r^2e^{ax}\sin(bx+c+2\theta)$

$$y_3=r^3e^{ax}\sin(bx+c+3\theta)$$

....

In general, $y_n = r^n e^{ax} \sin(bx+c+n\theta)$,

where $r=\sqrt{a^2+b^2}$ and $\theta = \tan^{-1}\dfrac{b}{a}$.

Similarly, if $y=e^{ax}\cos(bx+c)$, then
$$y_n = r^n e^{ax}\cos(bx+c+n\theta),$$
where r and θ have the same values as above.

Sometimes a given function can be transformed into one of the above standard forms and the nth derivative can then be found with the help of these formulae.

The following examples will illustrate the process.

Example 48. *If $y=\sin^2 x \cos^2 x$, find y_n.*

Solution.
$$y=\sin^2 x\cos^2 x=\tfrac{1}{4}\sin^2 2x$$
$$=\tfrac{1}{4}\cdot\tfrac{1}{2}(1-\cos 4x)$$
$$=\tfrac{1}{8}-\tfrac{1}{8}\cos 4x,$$

so that, $y_n=-\dfrac{1}{8}\cdot 4^n \cos\left(4x+\dfrac{n\pi}{2}\right)$.

Example 49. *If $y=e^{2x}\sin^3 x$, find y_n.*

Solution.
$$y=e^{2x}\sin^3 x$$
$$=e^{2x}\left[\tfrac{1}{4}(3\sin x-\sin 3x)\right]$$
$$=\tfrac{3}{4}e^{2x}\sin x-\tfrac{1}{4}e^{2x}\sin 3x.$$

Let $z=\tfrac{3}{4}e^{2x}\sin x,\ w=\tfrac{1}{4}e^{2x}\sin 3x.$

Then, $z_n=\tfrac{3}{4}(\sqrt{5})^n e^{x}\sin(x+n\tan^{-1}\tfrac{1}{2})$

and $w_n=\tfrac{1}{4}(\sqrt{13})^n e^{x}\sin(3x+n\tan^{-1}\tfrac{3}{2})$.

Since $y=z-w$,

therefore,
$$y_n=z_n-w_n=\tfrac{3}{4}(\sqrt{5})^n e^{x}\sin(x+n\tan^{-1}\tfrac{1}{2})$$
$$-\tfrac{1}{4}(\sqrt{13})^n e^{x}\sin(3\theta+n\tan^{-1}\tfrac{1}{2})$$

EXERCISE 3 (p)

Find the nth derivatives of

1. $(3x+4)^{n+2}$.
2. $(2x+1)^n$.
3. $\dfrac{1}{(2x+7)}$.
4. $\dfrac{1}{a-bx}$.
5. $\dfrac{1}{(3x+8)^2}$.
6. $\dfrac{x+5}{3x+4}$.
7. $\sqrt{(9x+8)}$.
8. $\ln(x+2)$.
9. 5^x.
10. $3e^{ax}$.
11. $\sin(3x+4)$.
12. $\cos(x+2)$.
13. $e^{3x}\sin 2x$.
14. $e^x\cos(x+5)$.
15. $\sin x\sin 3x$.
16. $\sin^2 x\cos^3 x$.
17. $\cos 4x$.
18. $\cos x\cos 2x\cos 3x$.
19. $e^x\cos 2x\cos 4x$.
20. $e^{3x}\sin x\sin 2x\sin 3x$.

TEST YOUR UNDERSTANDING III

In each of the following problems, four alternatives are given out of which one is correct. Put a tick-mark ($\checkmark$) against the correct alternative :

1. The derivative of $(3x^2+4)^3$ is
   (a) $3(3x^2+4)^2$ (b) $6x(3x^2+4)^2$
   (c) $18x(3x^2+4)^2$ (d) $9x^2(3x^2+4)$.
2. The derivative of $(\sin 2x^2)^3$ is
   (a) $3(\sin 2x^2)^2$ (b) $3(\sin 2x^2)^2 \cos 2x^2$
   (c) $12x(\sin 2x^2)^2 \cos 2x^2$ (d) $2(\sin 2x^2)^3 \cos 2x^2$.
3. The derivative of x^x is
   (a) x^{x-1} (b) $x^x \ln x$
   (c) $x^x \ln x$ (d) $x^x (1+\ln x)$.
4. The derivative of $\tan^{-1}(2x)$ is
   (a) $\dfrac{2}{\sqrt{(1-4x^2)}}$ (b) $\dfrac{-2}{\sqrt{(1-4x^2)}}$
   (c) $\dfrac{1}{1+4x^2}$ (d) $\dfrac{2}{1+4x^2}$
5. The derivative of e^{4x^2} is
   (a) e^{4x^2} (b) $8xe^{4x^2}$
   (c) $4x^2 e^{4x^2}$ (d) e^{4x^2-1}.
6. The derivative of $\sin^{-1}\dfrac{2x}{1+x^2}$ w.r.t. $\tan^{-1} x$ is
   (a) 2 (b) $\frac{1}{2}$
   (c) $\dfrac{1}{1+x^2}\sin^{-1}\left(\dfrac{2x}{1+x^2}\right)$ (d) $(1+x^2)\sin^{-1}\left(\dfrac{2x}{1+x^2}\right)$.
7. If $x=a(t+\sin t)$, $y=a(1-\cos t)$, then $\dfrac{dy}{dx}$ equals
   (a) $\tan \frac{1}{2}t$ (b) $\cot \frac{1}{2}t$
   (c) $\tan t$ (d) $-\cot t$.
8. The derivative of 2^{x^2} is
   (a) $2^{x^2} \ln 2$ (b) $x^2 . 2^{x^2} -1$
   (c) $2x . 2^{x^2} \ln 2$, (d) $2x\ 2^{x^2}$.
9. The derivative of $\ln(\sec x+\tan x)$ is
   (a) $\dfrac{1}{\sec x+\tan x}$ (b) $\sec x$
   (c) $\sec x-\tan x$ (d) $\tan(x/2)$.
10. The derivative of $\cos^{-1}(2x^2-1)$ is
    (a) $\dfrac{2}{\sqrt{(1-x^2)}}$ (b) $\dfrac{-2}{\sqrt{(1-x^2)}}$
    (c) $\dfrac{1}{1+4x^2}$ (d) $\dfrac{4x}{(2x^2-1)^2}$.

REVIEW EXERCISE III

1. Differentiate from first principles :

   (*a*) $ax^2+\dfrac{b}{x}\ (x\neq 0)$ (*A.I.S.S.C.E., 1986*)

   (*b*) $\sqrt{(x+1)}$ (*D.B.S.S.C.E., 1984*)

2. Differentiate from first principles :

   (*a*) $\sin 3x$ (*D.B.S.S.C.E., 1986*)

   (*b*) $\sin (2x+3)$ (*A.I.S.S.C.E., 1987*)

3. Differentiate from first principles :

   (*a*) $\cos 2x$ (*A.I.S.S.C.E., 1984*)

   (*b*) $\cos^2 x$ (*D.B.S.S.C.E., 1988*)

   (*c*) $\sqrt{\cos x}$ (*A.I.S.S.C.E., 1985*)

4. Differentiate from first principles :

   (*a*) $\tan 2x$ (*A.I.S.S.C.E., 1984*)

   (*b*) $\tan (3x+1)$ (*A.I.S.S.C.E., 1988*)

   (*c*) $\cot x$. (*A.I.S.S.C.E., 1986*)

5. Differentiate from first principles :

   (*a*) $\cot^{-1} x$ (*D.B.S.S.C.E., 1985*)

   (*b*) $\sin^{-1} x$ (*D.B.S.S.C.E., 1989*)

   (*c*) $\tan^{-1} x$. (*D.B.S.S.C.E., 1987*)

6. Differentiate e^{x^2} from first principles. (*A.I.S.S.C.E., 1989*)

7. Find the derivative of :

   (*a*) $\dfrac{\sqrt{x}\,(2x+3)^2}{\sqrt{x+1}}$ (*A.I.S.S.C.E., 1986*)

   (*b*) $\dfrac{\sqrt{x}\,(x+4)^{3/2}}{(4x-3)^{4/3}}$ (*A.I.S.S.C.E., 1988*)

8. Find the derivative of :

   (*a*) $\dfrac{1-\cos x}{1+\cos x}$ (*A.I.S.S.C.E., 1984*)

   (*b*) $e^{m\sin^{-1}x}\sin nx$ (*A.I.S.S.C.E., 1986*)

9. Differentiate with respect to x :

   (*a*) $e^x\, ln\,(1+x^2)$ (*A.I.S.S.C.E., 1987*)

   (*b*) $\sqrt{\{ln\,[\sin(\tfrac{1}{8}x^2-1)]\}}$ (*A.I.S.S.C.E., 1988*)

10. Find the derivative of :

    (*a*) x^x (*A.I.S.S.C.E., 1984*)

    (*b*) $(\sin x)^x$ (*D.B.S.S.C.E., 1987*)

11. Find $\dfrac{dy}{dx}$ if

    (*a*) $y=(\sin x)^{ln\,x}$ (*D.B.S.S.C.E., 1988*)

    (*b*) $y=x^{\sin x}+(\sin x)^x$ (*A.I.S.S.C.E., 1987*)

12. Find $\frac{dy}{dx}$ if

(a) $x^y = y^x$ (D.B.S.S.C.E., 1986)

(b) $(\tan^{-1} x)^y + y^{\cot x} = 1$.

13. Find $\frac{dy}{dx}$ when

(a) $y = \sin^{-1} \frac{2x}{1+x^2}$ (D.B.S.S.C.E., 1987)

(b) $y = \sec^{-1} \frac{1}{2x^2-1}$ (D.B.S.S.C.E., 1986)

14. Find $\frac{dy}{dx}$ if

(a) $y = \tan^{-1} \left(\frac{\cos x - \sin x}{\cos x + \sin x}\right)$ (A.I.S.S.C.E., 1984)

(b) $y = 4 \tan^{-1} \left(\frac{1+\cos 2x}{1-\cos 2x}\right)$ (A.I.S.S.C.E., 1985)

15. Find $\frac{dy}{dx}$ if

(a) $y = \tan^{-1} \sqrt{\frac{1-\cos x}{1+\cos x}}$ (A.I.S.S.C.E., 1984)

(b) $y = \tan^{-1} (\sec x + \tan x)$

(c) $y = \tan^{-1} \left(\frac{\sqrt{a} - \sqrt{x}}{1+\sqrt{ax}}\right)$

SUMMARY

1. *Algebra of derivatives.* Let f and g be two differentiable functions, and c a constant. Then

(a) $D[cf(x)] = c\, D[f(x)]$

(b) $D[(f+g)(x)] = D[f(x)] + D[g(x)]$

(c) $D[(fg)(x)] = D[f(x)]\, g(x) + f(x)\, D[g(x)]$

(d) $D\,[(f/g)(x)] = \frac{\{D[f(x)]\}\, g(x) - f(x)\, D[g(x)]}{[g(x)]^2}$,

provided $g(x) \neq 0$ for any x.

2. *Chain Rule for differentiation.* Let $y = g(u)$ and $u = f(x)$. If both $\frac{dy}{dx}$ and $\frac{du}{dx}$ exist, then $\frac{dy}{dx}$ exists and is given by

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

3. *Inverse function theorem for derivatives.* Let f and g be two strictly monotone functions which are inverses of each other. If f has at x_0 the derivative $f'(x_0) \neq 0$, then g has at $y_0\ [=f(x_0)]$ the derivative

$$g'(y_0) = \frac{1}{f'(y_0)}.$$

4. If $x = f(t)$, $y = g(t)$ be differentiable functions of t, then

$$\frac{dy}{dx} = \left(\frac{dy}{dt}\right) \Big/ \left(\frac{dx}{dt}\right).$$

5. Derivatives of Some Standard Functions

Function	*Derivative*
constant	0
x^r	rx^{r-1} (provided x^r, x^{r-1} are both defined)
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\tan x$	$\sec^2 x$
$\cot x$	$-\csc^2 x$
$\sec x$	$\sec x \tan x$
$\csc x$	$-\csc x \cot x$
$\sin^{-1} x$	$\dfrac{1}{\sqrt{(1-x^2)}}$
$\cos^{-1} x$	$\dfrac{-1}{\sqrt{(1-x^2)}}$
$\tan^{-1} x$	$\dfrac{1}{1+x^2}$
$\cot^{-1} x$	$-\dfrac{1}{1+x^2}$
$\sec^{-1} x$	$\dfrac{1}{\|x\| \sqrt{(x^2-1)}}$, if $\|x\|>1$
$\csc^{-1} x$	$\dfrac{-1}{\|x\| \sqrt{(x^2-1)}}$, if $\|x\|>1$
e^x	e^x
$\ln x$	$\dfrac{1}{x}$, if $x>0$
a^x	$a^x \ln a$, if $a>0$

6. Table of nth derivatives of some standard functions

Function	*nth derivative*
$(ax+b)^m$	$\dfrac{m!}{(m-n)!}\, a^n (ax+b)^{m-n}$, if $m \geqslant n$ and m is a positive integer
$\ln (ax+b)$	$(-1)^{n-1} \dfrac{(n-1)!}{x^n}$
a^{mx}	$m^n a^{mx} (\ln a)^n$
$\sin (ax+b)$	$a^n \sin\left(ax+b+\dfrac{n\pi}{2}\right)$
$\cos (ax+b)$	$a^n \cos\left(ax+b+\dfrac{n\pi}{2}\right)$
$e^{ax} \sin (bx+c)$	$r^n e^{ax} \sin (bx+c+n\theta)$, where $r=(a^2+b^2)^{\frac{1}{2}}$, $\theta=\tan^{-1}(b/a)$
$e^{ax} \cos (bx+c)$	$r^n e^{ax} \cos (bx+c+n\theta)$, where $r=(a^2+b^2)^{\frac{1}{2}}$, $\theta=\tan^{-1}(b/a)$.

HISTORICAL NOTE

Calculus was discovered by Issac Newton (1642-1727) and G.W. Leibnitz (1646-1717) independently in the second-half of the seventeenth century. Their work was preceded by contributions of many mathematicians. Fermat was the first to solve problems on maxima-minima by considering the behaviour of a function near the extreme values. Descartes devised a method for constructing tangent lines to a curve. Though algebraic in nature, his method exerted quite some influence on the development of the calculus. Formal algorithms for the constructions of tangents were discovered in the 1650's by the Dutch mathematicians Johann Hudde and Rene' Sluse. The work of Hudde and Sluse was followed by the discovery of infinitesimal methods for determination of tangents by Issac Barrow (1630-1677).

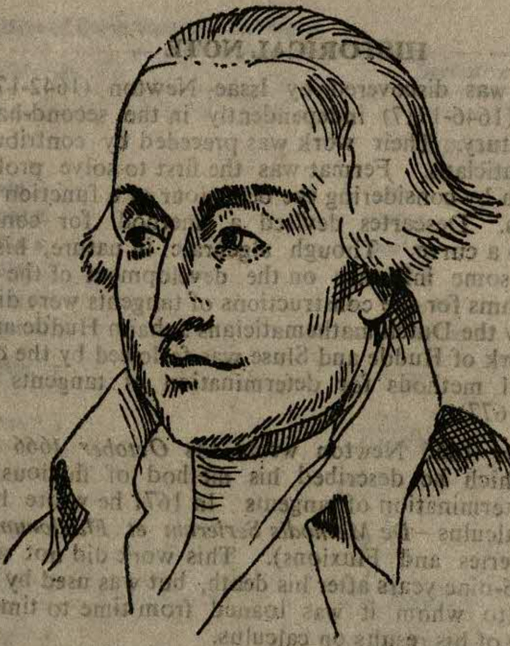
In October 1666 Newton wrote his *October 1666 Tract on Fluxions* in which he described his method of fluxious to treat problems of determination of tangents In 1671 he wrote his major treatise on Calculus—*De Methodis Serierum et Fluxionum* (of the Methods of Series and Fluxions). This work did not appear in print until 1736-nine years after his death, but was used by him and many others (to whom it was loaned from time to time) as the primary source of his results on calculus.

Leibnitz recorded his work on Calculus in a series of somewhat disjointed notes that he wrote during the last quarter of 1675. His first published article on differential calculus appeared in 1684 in the Leipzig periodical *Acta Eruditorum*. For Leibnitz, the separate differentials dx and dy were fundamental ; their ratio was "merely" a geometrically significant quotient.

Newton's formative work on the calculus dated from 1664 to 1966, while Leibnitz's work was done during 1672-1676. However, Leibnitz's work appeared in 1684 and 1686, whereas Newton did not publish anything on Calculus until his *Principia* of 1687 and his *Optiks* of 1704. The Newton-Leibnitz controversy as to who discovered calculus first, is one of the fiercest controversies regarding precedence, so far as mathematics is concerned. This unfortunate controversy has less to do with mathematics than the traditional nationalistic rivalry between English and continental European mathematicians.

The concept of rigour in calculus was introduced by A L. Cauchy, J.L. Lagrange and Karl Weierstress during the nineteenth century.

□□



JOSEPH LOUIS LAGRANGE (1736-1813)

Joseph Louis Lagrange, the greatest mathematician of the 18th century was born on January 25, 1736 of a French father and an Italian mother. During his life-time Lagrange was honoured by four royal personalities : Frederick the Great of Germany, the King of Sardinia, Napoleon Bonaparte (who called Lagrange the Lofty Pyramid of Mathematical Sciences) and Maria Antoinette of France.

Lagrange's first love were classics. He read Euclid and Archimedes, but they failed to win him over onto the side of mathematics. His conversion to mathematics was brought about by an essay of Halley on the superiority of the calculus methods over those of geometry.

At the age of 16, Lagrange was made the Professor of Mathematics at the Royal Artillery School in Turin. At the age of 19, he constructed the calculus of variations by means of which he unified mechanics in his great work Analytical Mechanics which Hamilton described as a scientific poem. He applied calculus to the theory of probability and proposed sound as a system of elastic particles. His memoir on the Solution of Numerical Equations was a source of inspiration from which drew all of Abel, Galois, Cauchy and others. His work on diophantine equations is noteworthy. The French Academy awarded him the Grand Prize for his solution of the 3-body problem. He won this Prize four times. It was he who devised the metric system of weights and measures.

Despite his mathematical distinctions, Lagrange was of an indifferent disposition. A perfect gentleman, he knew how to hold his tongue. He was often given to melancholy. He married twice, once early in life and once at the age of 56. His second wife was forty years younger than himself, but kept him well in tow and made him comfortable. He died in April 1813, well satisfied with life.

CHAPTER 4

Applications of the Derivative

4·1. INTRODUCTION

In the preceding chapter we have learnt as to how we can compute derivatives of functions. Derivatives have applications not only within mathematics, but also in physics, chemistry, biology, engineering, economics etc. In the present chapter we shall consider some simple applications of derivatives to motion in a straight line, motion under gravity, tangents and normals, determination of rate of change of quantities, determination of intervals in which a function is increasing or decreasing, maxima and minima of functions, and curve-sketching. We shall consider these applications one by one.

4·2. MOTION IN A STRAIGHT LINE

The derivative of a function expresses the rate of change of the function. However, before we discuss the concept of rate of change in general, let us consider the motion of a particle in a straight line. Without any loss of generality, we may think of a particle as just a point, and simply talk of a point moving in a straight line. In our discussion we shall use the words 'particle' and 'point' interchangeably while talking of motion in a straight line. Consider a straight line L (Fig. 4·1).

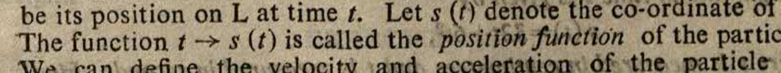


Fig. 4·1.

Let a point O on L be taken as the origin, A as the unit-point on L so that OA=1 unit. Let a particle be moving along L. Let P be its position on L at time t. Let $s(t)$ denote the co-ordinate of P. The function $t \to s(t)$ is called the *position function* of the particle. We can define the velocity and acceleration of the particle as follows :

Definition 4·1. *If s is the position-function of a particle, then its velocity at time t is defined as*

$$\lim_{h \to 0} \frac{s(t+h)-s(t)}{h},$$

and is denoted by $v(t)$.

Thus, $$v(t)=\frac{ds}{dt}.$$

Definition 4 2. *If $v(t)$ is the velocity of a particle moving in a straight line at time t, then its acceleration at time t is defined as*

$$\lim_{h\to 0}\frac{v(t+h)-v(t)}{h}$$

and is denoted by $a(t)$.

Thus, $$a(t)=\frac{dv}{dt}=\frac{d^2s}{dt^2}.$$

Remarks 1. Since we are dealing with a physical situation here, we assume that a particle in motion has a 'velocity' and an 'acceleration' at time t. Therefore we are assured that the limits in the above definitions exist.

2. Following Newton, the derivative of a function with respect to time is often denoted by a 'dot' overhead. Thus, instead of $\frac{ds}{dt}$ we often write $\dot{s}$. Similarly, $\frac{dv}{dt}$ is denoted by $\dot{v}$, and $\frac{d^2s}{dt^2}$ is denoted by $\ddot{s}$. We shall use this notation whenever convenient.

3. The relation between units of measurement of distance, time, velocity and acceleration is shown below :

Unit of distance	*Unit of time*	*Unit of velocity*	*Unit of acceleration*
cm	sec	cm/sec	cm/sec^2
m	min	m/min	m/min^2
km	hr	km/hr	km/hr^2

The following examples will illustrate the role of derivatives in discussing motion in a straight line.

Example 1. *The distance travelled by a particle moving in a straight line in time t is given by*

$$s(t)=32t+6t^2,$$

where s is measured in centimetres and t is measured in seconds. Find the velocity and acceleration at time t. Also find the initial velocity (i.e., velocity at $t=0$).

Solution. $s(t)=32t+6t^2$

$$v=\dot{s}=32+12t$$

$$\dot{v}=\ddot{s}=12.$$

Thus the velocity at time t is $32+12t$, initial velocity is 32 cm/sec, acceleration at time $t=12$ cm/sec^2.

Example 2. *The position of function of a particle moving in a straight line is given by $s(t)=3\cos 2t+4\sin 2t$. Find (a) the velocity at time t, (b) acceleration at time t (c) the maximum distance of the particle from the origin (d) the maximum velocity. Also show that the acceleration is directed towards the centre and is proportional to the distance of the particle from the origin.*

Solution. $s(t)=3\cos 2t+4\sin 2t$ ...(1)

Differentiating (1) twice throughout with respect to t, we have

$$\dot{s}=-6\sin 2t+8\cos 2t \quad ...(2)$$

$$\ddot{s}=-12\cos 2t-16\sin 2t \quad ...(3)$$

(*a*) Velocity at time t is given by (2) above.

(*b*) Acceleration at time t is given by (3) above.

(*c*) From (1) we have by putting

$3=r\cos\alpha,\ 4=r\sin\alpha$, (so that $r=5,\ \alpha=\tan^{-1}(4/3)$)

$$s(t)=r(\cos 2t\cos\alpha+\sin 2t\sin\alpha)$$
$$=r\cos(2t-\alpha)$$
$$=5\cos(2t-\alpha),\text{ where }\alpha=\tan^{-1}(4/3)). \quad ...(4)$$

Since $|\cos(2t-\alpha)|\leqslant 1$, therefore the maximum value of $s(t)$ is 5.

(*d*) As in (*c*) above, or by differentiating (4) throughout, we have

$$v(t)=-10\sin(2t-\alpha),$$

so that the maximum velocity is 10.

From (3), we find that

$$\ddot{s}=-4(3\cos 2t+4\sin 2t)=-4s,$$

showing that the acceleration is proportional to the distance travelled. The negative sign shows that acceleration is directed towards the origin.

EXERCISE 4 (*a*)

1. The distance travelled in t seconds by a particle moving in a straight line is given (in metres) by $s(t)$.

   Find the velocity and acceleration when

   (*a*) $s(t)=2t^2,\ t=2$.

   (*b*) $s(t)=10t+\frac{1}{4}t^2,\ t=3$.

   (*c*) $s(t)=20t-4t^2,\ t=2{\cdot}5$.

2. A particle is moving in a straight line so that its position at time t is given by

$$s(t)=\mathrm{A}t^2+\mathrm{B}t+\mathrm{C},$$

A, B, C being constants.

If at time 2 seconds, it is at a distance 8m from the origin, its velocity is 3m/sec, and acceleration is 2m/sec², find A, B, C. Also find its position at time $t=3$ seconds.

3. A particle moves in a straight line. Its distance from a fixed point on the line at time t seconds is $2t^2+6t+1$ metres. Find its velocity and acceleration where $t=2{\cdot}5$ sec.

4. The distance $s(t)$ travelled by a particle moving in a straight line is measured in metres, and is given by

$$s(t)=2t^3-9t^2+12t+6,$$

where t is measured in seconds.

Find where its acceleration becomes zero and the velocity at that instant.

5. A particle moves in a fixed straight line so that $s(t)=\sqrt{t}$. Show that the acceleration is negative and proportional to the cube of the velocity.

6. A particle moves in a straight line so that its distance $s(t)$ from a fixed point on the line at any instant t is $3t^n$. If v is the velocity and f the acceleration at time t, show that

$$v^2=nfs/(n-1).$$

7. A particle moves along a straight line such that its distance x from a fixed point on it and the velocity v are related by

$$v^2=36\,(9-x^2).$$

Show that the acceleration varies as the distance of the particle from the origin and is directed towards the origin.

8. A train starting at time $t=0$ moves in time t, a distance $y=100t(1-e^{-2t})$. Find the velocity and acceleration at time t.

4·3. MOTION UNDER GRAVITY

Galileo discovered in 1589 that when a body falls vertically near the surface of the earth (air resistance being neglected), it does so with a constant acceleration. In his famous experiments conducted while he was at Pisa, he showed that all bodies near the earth's surface fall with the same acceleration, whatever their masses.

In the early part of the seventeenth century Johann Kepler (1571-1630) put forth his three laws of planetary motion which were based on the observational data collected by the Czech astronomer Tycho Brahe.

During the later part of the seventeenth century Newton propounded his universal law of gravitation and three laws of motion. He showed that Kepler's empirical laws could be deduced as a consequence of his laws, thus providing a justification for the universal law of gravitation. According to this law, *two particles of masses m and m', at a distance r apart, attract each other with a force Gmm'/r^2, where G is a universal constant depending only upon the system of units employed.* It can be shown that as a consequence of this law the mutual force of attraction between two spheres of masses m and m' is Gmm'/r^2 where r is the distance between their centres. This implies that near the surface of the earth, the gravitational force on a body of mass m is $m\left(\frac{GM}{R^2}\right)$, where M is the mass of the earth, and R is the radius of the earth. This force, called the weight of the body, is approximately 9 8 m newtons, if the mass is measured in kilograms. If the body is allowed to fall freely, it will have the acceleration GM/R^2 which is independent of the mass of the body. Since the earth is not exactly a sphere, the acceleration, usually denoted by g, varies in magnitude from 9 78 m/sec^2 at the equator to 9·83 m/sec^2 at the North Pole. We shall, however, take the value of g to be constant, equal to 9·8 m/sec^2. (The magnitude of this acceleration also varies with height, but the variation is rather small). The direction of this acceleration is 'towards the centre of the earth'. In fact, it defines 'the vertically downwards direction'. This is not strictly accurate because of the effect of earth's rotation, which again we neglect.

4 3.1. Motion in a Vertical straight line

If a particle is thrown vertically upwards with an initial velocity u, it will move with an acceleration g (directed downwards). The position function $s(t)$ is given by

$$s(t)=ut-\tfrac{1}{2}gt^2.$$

If a particle is let fall from a point above the ground, the position-function is given by

$$s(t)=\tfrac{1}{2}gt^2.$$

We have already considered such motion in the preceding section.

4·3 2. The path of a projectile

A body projected from a point, into the air, in a direction other than the vertical, describes a curved path. A body so projected is called a *projectile* and the path described by it is called its trajectory. The motion of a projectile in actual practice is quite a complex one. However, if we assume that (*i*) the body projected is so small that it can be regarded as a particle, (*ii*) the air-resistance is neglected, (*iii*) acceleration due to gravity is assumed to be

constant throughout the motion, then the path of a projectile is a parabola.

Let O be the point of projection of a particle and let OX and OY be a pair of rectangular axes lying in the plane of projection with OX horizontal, and OY vertical.

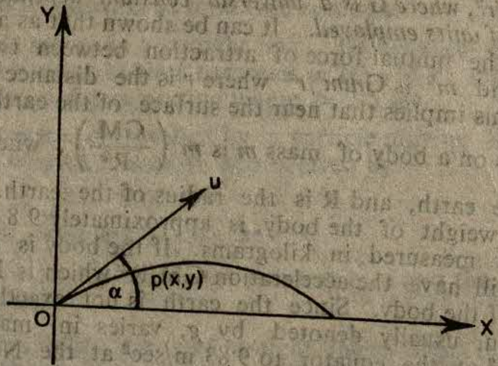


Fig. 4·2.

If u be the velocity of projection, α the angle which the direction of projection makes with OX, and $P(x, y)$ the position of the particle at time t, then the motion is given by

$$\left.\begin{aligned} x&=(u\cos\alpha)t \\ y&=(u\sin\alpha)t-\tfrac{1}{2}gt^2 \end{aligned}\right\} \quad \text{...(1)}$$

Differentiating the above relations with respect to t, we have

$$\left.\begin{aligned} \dot{x}&=u\cos\alpha \\ \dot{y}&=u\sin\alpha-gt \end{aligned}\right\} \quad \text{...(2)}$$

as the components of the velocity of the particle at time t, along OX and OY.

The above equations show that the horizontal component of the velocity ($\dot{x}$) is constant, and the verticle component ($\dot{y}$) is the same as if the particle were projected vertically upwards with a velocity $u \sin \alpha$. Differentiating (2) throughout with respect to t, we have

$$\left.\begin{aligned} \ddot{x}&=0 \\ \ddot{y}&=-g \end{aligned}\right\} \quad \text{...(3)}$$

Equations (3) give the acceleration of the particle at time t (in fact, at any time—because the acceleration is constant, being equal to g vertically downwards).

Remark. Equations (3) are a consequence of Newton's second law of motion. Equations (1) and (2) are obtained by *integrating* equations (3). You will learn the meaning and techniques of integration in the next two chapters. Here we have only tried to give you a glimpse of an important application of calculus.

Example 3. *A cricket ball is projected at an angle $\pi/4$ to the horizontal with a velocity 24·5 m/sec. Find the equation of the path and the maximum height which it attains during the motion (Take $g=9{\cdot}8$ m/sec^2).*

Solution. The position of the ball at time t is given by

$$x=24{\cdot}5\cos(\pi/4)t=\frac{49}{2\sqrt{2}}\,t, \qquad \cdots(1)$$

$$y=24{\cdot}5\sin(\pi/4)\,t-\frac{1}{2}\,(9{\cdot}8)t^2$$

$$=\frac{49}{2\sqrt{2}}\,t-4{\cdot}9t^2 \qquad \cdots(2)$$

Eliminating t from (1) and (2), we have

$$y=x-4{\cdot}9\left(\frac{2\sqrt{2}}{49}\,x\right)^2=x-\frac{4}{245}\,x^2,$$

as the equation of the path.

Differentiating (2) with respect to t, we have

$$\dot{y}=\frac{49}{2\sqrt{2}}-9{\cdot}8\,t$$

The ball stops rising above when $\dot{y}=0$.

Now $\dot{y}=0$, when $t=\dfrac{49}{2\sqrt{2}\times 9{\cdot}8}=\dfrac{5}{2\sqrt{2}}$.

When $t=\dfrac{5}{2\sqrt{2}}$, $y=\dfrac{49}{2\sqrt{2}}\left(\dfrac{5}{2\sqrt{2}}\right)-4{\cdot}9\left(\dfrac{5}{2\sqrt{2}}\right)^2$

$$=\frac{245}{16}\text{ metres.}$$

EXERCISE 4 (*b*)

1. The distance travelled in t seconds by a ball dropped vertically downwards from a tower is given by $s=4{\cdot}9t^2$. What will be its velocity when it has travelled for 3 seconds ?
2. A ball is dropped vertically downwards from the top of a tower 40 meters high. The distance travelled by the ball in time t is given by $s=4{\cdot}9t^2$. What will be the velocity of the ball when it reaches the ground ?

3. A cricket ball is thrown from the ground vertically upwards. Its height at time t seconds is given by $s(t)=14t-4 \cdot 9t^2$. What will be its velocity after 2 seconds? How high will the ball go?

4. A particle is projected with a velocity 7 m/sec. at an angle $60°$ to the horizontal. Find the equation of its path. How much time does it take to reach highest point?

5. A ball is projected at an angle $30°$ to the horizontal. The maximum height attained by it is 20 metres. Find the velocity of projection.

6. A particle is projected at an angle $30°$ to the horizontal and returns to the ground in 10 seconds. Find the velocity of projection.

7. A particle is projected at an angle $60°$ to the horizontal. If the velocity of projection is 14 m/sec, find the greatest height attained by it.

4·4. RATE OF CHANGE OF QUANTITIES

We have seen that if s be the distance travelled by a particle, then ds/dt, the rate of change of distance travelled with respect to time, represents the velocity of the particle. Similarly dv/dt, the rate of change of velocity with respect to time, represents the acceleration of the particle. More generally, if a quantity y varies with respect to another quantity x according to the functional relation $y=f(x)$, then the derivative $f'(c)$ represents the rate of change of y with respect to x, when $x=c$. Furthermore, if two quantities x and y are related to each other by the functional relation $y=f(x)$, and x and y both vary with t, so that x and y are both given functions of t, then we have

$$\frac{dy}{dt}=\frac{dy}{dx} \cdot \frac{dx}{dt},$$

i.e.,

$$\frac{dy}{dt}=f'(x)\frac{dx}{dt}. \qquad ...(1)$$

From (1) we find that if we are given f and dx/dt, we can find dy/dt. The following example will illustrate how we can use (1) to find the rate of change of a quantity when the rate of change of a related quantity is given.

Example 4. *A spherical soap bubble is expanding so that its radius is increasing at the rate of 0·02 cm per second. At what rate is surface area increasing when its radius is 4 cm? (Take $\pi=3 \cdot 14$).*

Solution. Let the surface area of the bubble be S cm² when its ratius is r cm. Then the relation between S and r is given by

$$S=4\pi r^2 \qquad (1)$$

Differentiating both sides of (1) with respect to t we have,

$$\frac{dS}{dt}=8\pi r\frac{dr}{dt} \qquad ...(2)$$

We are given that $\frac{dr}{dt}=\cdot 02$ when $r=4$. Therefore from (2), we have

$$\frac{dS}{dt}=8\times 3\cdot 14\times 4\times \cdot 02=2\cdot 0096,$$

so that the surface area is increasing at the rate of 2·0096 cm^2/sec.

Remark. Following Newton, we can denote derivative with respect to time by a dot, so that $\frac{dr}{dt}$, $\frac{dS}{dt}$ can be denoted by $\dot{r}$ and $\dot{S}$ respectively. We shall often use this notation.

Example 5. *Two buses start from a certain place at the same instant. One goes east at 45 km per hour, and the other goes north at 60 km per hour. How fast is the distance between them increasing 3 hours later ?*

Solution. Let the buses start from O, one of them going eastwards along OX, and the other going northwards along OY. At time t, let A and B denote the positions of the two buses. If OA$=x$, OB$=y$, AB$=d$, then from the right-angled triangle OAB, we have

$$d^2=x^2+y^2. \qquad ...(1)$$

Hence x, y and d are all functions of t.

Differentiating both sides of (1) with respect to t, and denoting derivatives with respect to t by dots, we have

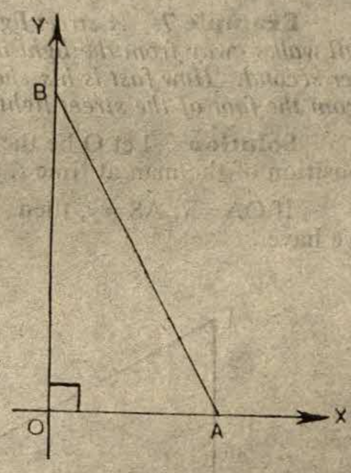


Fig 4.3

$$2d\,\dot{d}=2x\,\dot{x}+2y\,\dot{y},$$

or $$d\,\dot{d}=x\,\dot{x}+y\,\dot{y}. \qquad ...(2)$$

We have to find $\dot{d}$ when $t=3$.

We are given that $\dot{x}=45$, $\dot{y}=60$, $x=3.45=135$, $y=3.60=180$, $d=\sqrt{(x^2+y^2)}=\sqrt{\{(3.45)^2+(4.45)^2\}}=225$. Substituting these values in (2), we have

$$\begin{aligned}225\dot{d}&=135.45+180.60,\\&=225(27+48),\\&=225.75,\end{aligned}$$

or $$\dot{d}=75.$$

Thus the distance between the buses is increasing at the rate of 75 km per hour.

Example 6. *The pressure (p) and the volume (v) of a given mass of gas are connected by the relation pv=200. Find the rate of change of volume when the pressure is 8 gm/cm² per second and it is increasing at the rate of ·4 gm/cm².*

Solution. We are given that

$$pv=200 \qquad ...(1)$$

Differentiating both sides with respect to t, we have

$$\dot{p}\, v+p\, \dot{v}=0. \qquad ...(2)$$

When $p=8$, from (1) we have $v=25$.

Also, $p\ \dot{p}=\cdot 4$. Therefore from (2), we have

$$\cdot 4\times 25+8\dot{v}=0,$$

or

$$\dot{v}=-1\cdot 25.$$

Therefore the volume is decreasing at the rate of $1\cdot 25\ cm^3$ per second.

Example 7. *A street light is 5m above the ground. A man 2m tall walks away from the light in a straight line at the rate of 1·5 m per second. How fast is his shadow lengthening when he is 40m away from the foot of the street light ?*

Solution. Let O be the foot of the light, L the light, AB the position of the man at time t, and AS the shadow at time t.

If $OA=x$, $AS=y$, then from similar triangles ABS and OLS, we have

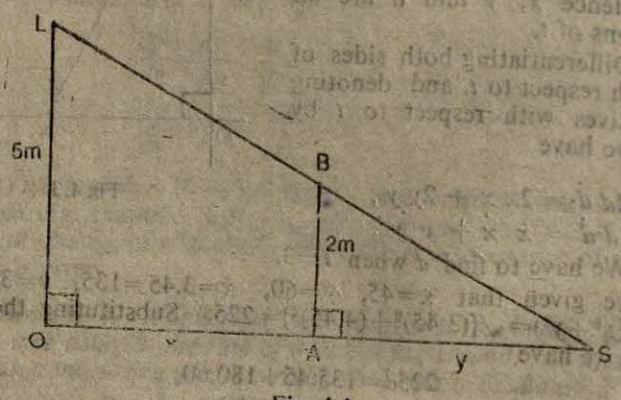


Fig. 4·4.

$$\frac{y}{x+y}=\frac{2}{5}, \text{ so that } 5y=2(x+y), \text{ i.e., } y=\frac{2}{3}x.$$

Differentiating both sides of the relation $y=\frac{2}{3}x$ with respect to t, we have

$$\dot{y}=\tfrac{2}{3}\dot{x}$$

Since $\dot{x}=1{\cdot}5$, therefore $\dot{y}=\frac{2}{3}\times 1{\cdot}5=1$.

Therefore the length of the shadow is increasing at 1m per sec.

Remark. Observe that $\dot{y}$ is independent of the distance of the man from the street light.

EXERCISE 4 (*c*)

1. Water is flowing into a cyclindrical tank of radius 60 cm at the rate of 30000 cm^3/min. How fast is the water level rising ?
2. The surface area of a spherical bubble is increasing at 2 cm^2/sec. When the radius of the bubble is 6 cm, at what rate is the volume of the bubble increasing ?
3. The height of a cone is 25 cm. If the radius of its base increases at ·5 cm/min and its height remains constant, at what rate is its volume increasing at the instant when the radius of its base is 30 cm ?
4. An aeroplane at an altitude 800 m flying horizontally at 720 km/hr passes directly over an observer. At what rate is it approaching the observer when it is 1000 m away from him ?
5. A sphere is expanding so that its surface is increasing at the rate of 0·01 cm^2/sec. Find the rate of increase of its volume when its radius is 10 cm.
6. A ladder 6 m in length is resting against a vertical wall. The bottom of the ladder is pulled along the ground away from the wall, at the rate of 1·5 cm/sec. How fast is the height of the highest point of the ladder decreasing when the foot of the ladder is 4·8 m away from the wall ?
7. Sand is poured at the rate of 20 cm^3/sec so as to form a conical pile whose height is always one-third of the radius of the base. At what rate is the radius of the base increasing when its height is 10 cm ?
8. A rod AB, 5 m long, moves with its ends A and B on two perpendicular lines OX and OY respectively. If A is 3 m from O and is moving away from it at the rate of 1 m/sec, find at what rate the end B is moving.
9. An inverted cone has a depth 10 cm and the radius of its base is 5 cm. Water is poured into it at the rate of 1 cm^3/min. At what rate is the water level in the cone rising when the depth is 3 cm ?

10. A vessel containing water is in the form of an inverted hollow cone with semi-vertical angle 30°. Water is running out of the cone at the rate of 4 cm^3/sec. through a small hole at the vertex of the cone. At what rate is the surface area in contact with water changing when there is 648π cm^3 of water remaining in the cone ?

4·5. DIFFERENTIALS ; ERRORS AND APPROXIMATIONS

Suppose we are given a function f, and c is a point in the domain of f. Then

$$f'(c)=\lim_{\Delta x\to 0}\frac{f(c+\Delta x)-f(c)}{\Delta x}. \qquad ...(1)$$

If we write $y=f(x)$, and denote by Δy the change in y as x changes from c to $c+\Delta x$, then

$$\Delta y=f(c+\Delta x)-f(c), \qquad ...(2)$$

and (1) may be written as

$$f'(c)=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} \qquad ...(3)$$

The right hand side is usually denoted by $\frac{dy}{dx}$, as we have done all along in the preceding chapter. Leibnitz, while developing calculus, thought of $\frac{dy}{dx}$ as the ratio of dy and dx. He called dy and dx as *differentials*. *If we do that*, we can rewrite (2) as

$$dy=f'(c)\ dx. \qquad ...(4)$$

In (3), we can think of dx as one variable, and dy as another variable defined by the relation (4). If we take $\Delta x=dx$, then (2) will become

$$\Delta y=f(c+dx)-f(c) \qquad ...(5)$$

Geometrically, the quantities Δx, Δy, dy are related in the manner as shown in Fig. 4·5.

As an illustration, consider the function

$$f(x)=x^3+3 \text{ and } c=2$$

Then $\Delta y=(2+\Delta x)^3+3-(2^3+3)=12\Delta x+6(\Delta x)^2+(\Delta x)^3$.

Letting $\Delta x=dx$, we have

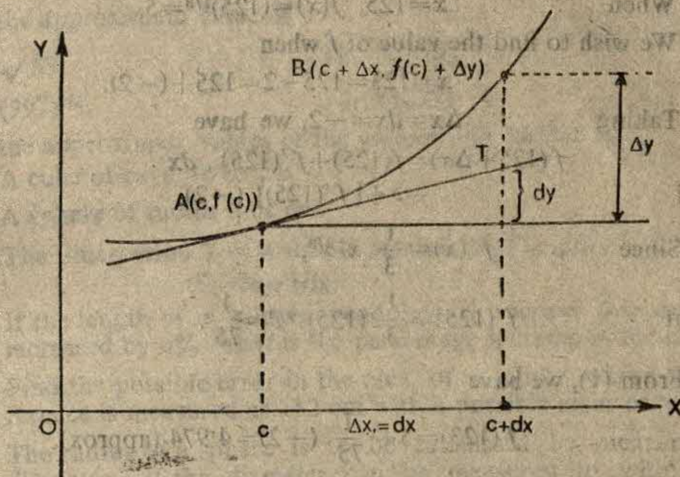


Fig. 4·5.

$$\Delta y = 12\, dx + 6(dx)^2 + (dx)^3.$$

Also, $\quad dy = f'(2)dx = 12\, dx.$

Observe that Δy and dy are approximately equal; they differ in second and higher powers of dx. As dx takes smaller and smaller values, this difference becomes smaller and smaller. For example,

If $dx = \cdot 1$, then in the above illustration

$\Delta y = 1\cdot 261,\ dy = 1\cdot 2.$

If $dx = \cdot 01$,

$\Delta y = \cdot 120601,\ dy = \cdot 12.$

The fact that when dx is small, the differential dy is an approximation to Δy is often used to calculate errors. The following examples will illustrate the method.

Example 8. *Find dy for the function given by*

$$y = x^4 - 3x^3 + 6x + 7, \text{ when } x = 2 \text{ and } dx = \cdot 01.$$

Solution. Let $f(x) = x^4 - 3x^2 + 6x + 7$,

so that $\quad f'(x) = 4x^3 - 6x + 6$

Now $\quad dy = f'(x)\, dx,$

$\quad = (4x^3 - 6x + 6)\, dx \quad \ldots(1)$

When $\quad x = 2$ and $dx = \cdot 01$, (1) yields

$dy = (4.2^3 - 6.2 + 6) \times \cdot 01$

$\quad = 26 \times \cdot 01 = \cdot 26.$

Example 9. *Find the approximate value of* $(123)^{1/3}$.

Solution. Let $f(x) = x^{1/3}$.

When $x=125,\ f(x)=(125)^{1/3}=5.$

We wish to find the value of f when

$$x=123=125-2=125+(-2).$$

Taking $\Delta x=dx=-2$, we have

$$f(125+\Delta x)=f(125)+f'(125)\,.\,dx$$
$$=5+[f'(125)]\,(-2) \qquad ...(1)$$

Since $f'(x)=\frac{1}{3}\,x^{-2/3}$,

so that $f'(125)=\frac{1}{3}\,(125)^{-2/3}=\frac{1}{75}.$

From (1), we have

$$f(123)=5+\frac{1}{75}\,(-2)=4{\cdot}974 \text{ (approx)}.$$

Example 10. *The radius of a sphere has been found to be 10 cm by measurement. If there is a possible error of ·01 cm in the measurement, find the possible error in the calculated volume.*

(*Take* $\pi=3{\cdot}14$).

Solution. If r be the radius of the sphere and V be its volume, then we know that

$$V=\frac{4}{3}\,\pi r^3 \qquad ...(1)$$

Taking differentials of both sides of (1), we have

$$dV=4\pi r^2\,dr,$$
$$=4\times 3{\cdot}14\times 100\times {\cdot}01 \text{ cm},$$
$$=12{\cdot}56 \text{ cm}^3.$$

EXERCISE 4 (*d*)

Find dy in terms of x and dx :

1. $y=x^3-3x+1.$
2. $y=\dfrac{x-2}{x^2+2}.$
3. $y=(2x+1)^{5/2}.$
4. $y=\sin 3x \cos 2x.$
5. $y=e^{2x}\ln(x-1).$
6. $y=e^{-3x}\cos 2x.$

Find dy for the values given for x and dx :

7. $y=x^2+2x-1,\ x=2,\ dx={\cdot}01.$
8. $y=x^3-6x+2,\ x=3,\ dx={\cdot}02.$
9. $y=\sqrt{x},\ x=1,\ dx={\cdot}04.$
10. $y=\sin 2x,\ x=\pi/6,\ dx={\cdot}03.$

Find the approximate value of :

11. $\sqrt{63}$. 12. $^3\sqrt{126}$.

13. $(997)^{1/3}$. 14. $(1028)^{1/10}$.

Find the approximate values of the volume and surface of.

15. A cube of edge 10·03 cm.

16. A sphere of radius 6·02 cm.

17. The time-period T of a simple pendulum of length l is given by
$$T=2\pi\sqrt{(l/g)}.$$
If the length of a simple pendulum of period 4 seconds is increased by 4%, what is the percentage increase in the period?

18. Find the possible error in the area of a circle whose circumference is measured as 112 cm with a possible error of ·05 cm.

19. The radius of a sphere is to be calculated by measuring its diameter. If the diameter can be measured to within 0·2, what is the maximum percentage error in the volume calculated ?

20. A cubical box has an edge of length 10 cm with a possible error of ·1 cm. Find the possible error in its volume.

4·6. TANGENTS AND NORMALS

In this section we shall use derivatives to find the equations of the tangent and normal to a given curve.

4·6·1. Equation of the Tangent

We know that if ψ is the angle which the tangent at any point (x, y) on the curve $y=f(x)$ makes with the x-axis, then
$$\tan\psi=\frac{dy}{dx}=f'(x).$$

Therefore, the equation of the tangent at the point (x, y) on the curve $y=f(x)$ is
$$Y-y=f'(x)(X-x),$$
where (X, Y) are the current co-ordinates of any point on the tangent.

Tangent to the curve x=f(t), y=f(t), at the point 't'.

At any point 't' of the curve $x=f(t)$, $y=g(t)$, at which $f'(t)\neq 0$, the slope of the tangent is
$$\frac{dy}{dx}=\frac{dy}{dt}\,\frac{dt}{dx}=\frac{g'(t)}{f'(t)}.$$

Hence the equation of the tangent to the curve
$$x=f(t), y=g(t)$$

at the point 't' is

$$[Y-g(t)]f'(t)=[X-f(t)]g'(t).$$

Example 11. *Find the equation of the tangent to the curve $y=2x^3-x^2+3$ at the point $(1, 4)$.*

Solution. Differentiating the relation

$$y=2x^3-x^2+3$$

throughout with respect to x, we have

$$\frac{dy}{dx}=6x^2-2x.$$

At the point $(1, 4)$, $\dfrac{dy}{dx}=6.1^2-2.1=4.$

$\therefore$ The slope of the tangent at $(1, 4)=4$.

Hence the equation of the tangent at $(1, 4)$ is

$$y-4=4(x-1),$$

or $$y=4x.$$

Example 12. *Find the equation of the tangent to the cycloid $x=a(\theta+\sin\theta)$, $y=a(1-\cos\theta)$ at the point 'θ'.*

Solution. We have

$$\frac{dx}{d\theta}=a(1+\cos\theta),\quad \frac{dy}{d\theta}=a\sin\theta.$$

$$\therefore\quad \frac{dy}{dx}=\frac{dy}{d\theta}\times\frac{d\theta}{dx},$$

$$=\frac{dy}{d\theta}\bigg/\frac{dx}{d\theta}$$

$$=\frac{\sin\theta}{1+\cos\theta},\ \text{except when } \cos\theta=-1.$$

$$=\tan(\theta/2).$$

$\therefore$ The slope of the tangent at the point θ is $\tan(\theta/2)$.

Hence the equation of the tangent at the point 'θ' is

$$y-a(1-\cos\theta)=\tan(\theta/2)\{x-a(\theta+\sin\theta)\},$$

or $$x\sin(\theta/2)-y\cos(\theta/2)$$
$$=a(\theta+\sin\theta)\sin(\theta/2)-a(1-\cos\theta)\cos(\theta/2),$$

or $$x\sin(\theta/2)-y\cos(\theta/2)=a\theta\sin(\theta/2).$$

Example 13. *Prove that the straight line $x/a+y/b=2$ touches the curve $(x/a)^n+(y/b)^n=2$ at the point (a, b), whatever be the value of n.*

Solution. Since the point (a, b) lies on the curve

$$(x/a)^n+(y/b)^n=2, \quad ...(1)$$

as well as on the line

$$x/a+y/b=2, \quad ...(2)$$

we have only to show that the slope of the tangent to (1) at the point (a, b) is $-b/a$ [the slope of the line (2)], whatever be the value of n.

Differentiating (1) throughout with respect to x, we have

$$n(x/a)^{n-1} \cdot \frac{1}{a} + n(y/b)^{n-1} \cdot \frac{1}{b} \cdot \frac{dy}{dx} = 0,$$

or
$$\frac{dy}{dx} = -\frac{b}{a} \cdot (x/a)^{n-1}/(y/b)^{n-1}$$

$\therefore$ The slope of the tangent at (a, b) is

$$-(b/a)\,(a/a)^{n-1}/(b/b)^{n-1} = -b/a.$$

whatever be the value of n.

Hence the curve (1) touches the line (2) at the point (a, b), whatever be the value of n.

EXERCISE 4 (*e*)

1. Find the equation of the tangent to each of the following curves at the point (x', y') :
   (*i*) $x^2+y^2+2gx+2fy+c=0$
   (*ii*) $x^2/a^2+y^2/b^2=1.$
   (*iii*) $x^2/a^2-y^2/b^2=1.$
   (*iv*) $xy=c^2.$
   (*v*) $y^2=4ax.$
2. Find the equation of the tangent to each of the following curves at the point 'θ' :
   (*i*) $x=a \cos \theta,\ y=b \sin \theta.$
   (*ii*) $x=a \sec \theta,\ y=b \tan \theta.$
   (*iii*) $x=a \cos^3 \theta, y=a \sin^3 \theta.$
   (*iv*) $x=a(\theta-\sin \theta),\ y=a(1-\cos \theta).$
3. Find the equation of the tangent to the curve $3xy^2-2x^2y=1$ at the point $(1, 1)$.
4. Find the points on curve $y=(x-1)(x-2)(x-3)$ at which the tangents are parallel to the axis of x.
5. Prove that the line $x/a+y/b=1$ touches the curve $y=be^{-x/a}$ at the point where it crosses the axis of y.
6. Find the equations of the tangents to the curve $y=x^3$ parallel to the line $y=12x+1$.
7. Find the condition that the line $x \cos \alpha+y \sin \alpha=p$ may touch the ellipse $x^2/a^2+y^2/b^2=1.$

8. If the line $x \cos \alpha + y \sin \alpha = p$ touches the curve
$$\left(\frac{x}{a}\right)^{n/(n-1)} + \left(\frac{y}{b}\right)^{n/(n-1)} = 1,$$
prove that $(a \cos \alpha)^n + (b \sin \alpha)^n = p^n$.

9. Show that the line $x \cos \alpha + y \sin \alpha = p$ touches the curve $x^m y^n = a^{m+n}$, provided
$$p^{m+n} m^m n^n = (m+n)^{m+n} a^{m+n} \cos^m \alpha \sin^n \alpha.$$

10. Prove that all points of the curve $y^2 = 4a\{x + a \sin (x/a)\}$ at which the tangent is parallel to the axis of x lie on a parabola.

4·6·2. Equations of the normal at a point

The **normal** to a curve at a point is the straight line which passes through the point and is perpendicular to the tangent to the curve at the point.

To obtain the equation of the normal to the curve $y = f(x)$ at any point (x, y), we first find the slope of the normal. If m be the slope of the normal, then
$$m \frac{dy}{dx} = -1,$$
so that
$$m = -1 \bigg/ \frac{dy}{dx}.$$

The normal being the line through (x, y) with slope m, its equation is
$$Y - y = \left(\frac{-1}{\dfrac{dy}{dx}}\right)(X - x),$$
or
$$\frac{dy}{dx}(Y - y) + (X - x) = 0, \qquad \ldots(1)$$
where (X, Y) are the current co-ordinates of a point.

Hence the equation of the normal at (x, y) to the curve $y = f(x)$ is given by (1).

The equation of the normal at any point 't' on the curve $x = f(t)$, $y = g(t)$ is given by
$$(X - f(t))f'(t) + (Y - g(t))g'(t) = 0.$$

Example 14. *Find the equation of the normal to the hyperbola*
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$
at the point $(a\sqrt{2}, b)$.

Solution. Differentiating the relation

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$$

throughout with respect to x, we have

$$\frac{2x}{a^2}-\frac{2y}{b^2}\frac{dy}{dx}=0,$$

or $$\frac{dy}{dx}=\frac{b^2x}{a^2y}.$$

$\therefore$ The slope of the tangent at $(a\sqrt{2}, b)$ is

$$\frac{b^2.a\sqrt{2}}{a^2b}=\frac{b\sqrt{2}}{a}$$

The slope 'm' of the normal at $(a\sqrt{2}, b)$ is, therefore, given by

$$m\frac{b\sqrt{2}}{a}=-1,$$

or $$m=-\frac{a}{b\sqrt{2}}.$$

The equation of the normal at $(a\sqrt{2}, b)$ is

$$y-b=-\frac{a}{b\sqrt{2}}(x-a\sqrt{2}),$$

or $$ax+b\sqrt{2}y=(a^2+b)\sqrt{2}.$$

Example 15. *Find the equation of the normal to the curve $y=x^3$ which is parallel to the line*

$$x+12y=1.$$

Solution. Differentiating the relation $y=x^3$ throughout with respect to x, we have

$$\frac{dy}{dx}=3x^2.$$

The slope of the normal at (x, y) is

$$\left(\frac{-1}{\frac{dy}{dx}}\right)=-\frac{1}{3x^2}$$

The normal is parallel to the line $x+12y=1$, provided

$$-\frac{1}{3x^2}=-\frac{1}{12},$$

i.e., $$x=\pm 2.$$

The points on the given curve whose abscissae are ± 2 are $(2, 2^3)$ and $(-2. (-2)^3)$, *i.e.*, $(2, 8)$ and $(-2, -8)$.

The normal at $(2, 8)$ is the line through this point parallel to

$$x+12y=1,$$

i.e., $$(x+2)+12(y-8)=0,$$

i.e., $$x+12y-98=0. \qquad \text{...(1)}$$

Similarly, the equation of the normal at $(-2, -8)$, is

$$(x+2)+12(y+8)=0,$$

i.e., $$x+12y+98=0. \qquad \text{...(2)}$$

From (1) and (2), we find that the given curve has two normals parallel to $x+12y=1$, and their equations are

$$x+12y \pm 98=0.$$

EXERCISE 4 (*f*)

1. Find the equation of the normal at (x', y') to each of the following curves :
   (*i*) $x^2+y^2=r^2$.
   (*ii*) $y^2=4ax$.
   (*iii*) $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$.
   (*iv*) $\dfrac{x^2}{a^2}-\dfrac{y^2}{b^2}=1$.
2. Find the equation of the normal at the point 'θ' to each of the following curves :
   (*i*) $x=a\cos\theta,\ y=b\sin\theta$.
   (*ii*) $x=a\sec\theta,\ y=b\tan\theta$.
   (*iii*) $x=a(\theta+\sin\theta),\ y=a(1-\cos\theta)$.
   (*iv*) $x=a\cos^3\theta,\ y=a\sin^3\theta$.
3. Find the equation of the normal to the curve
   $$y(x-2)(x-3)-x+7=0$$
   at the point where it cuts the axis of x.
4. Find the equation of the normal at (2, 2) to the curve
   $$x^2y^3=32.$$
5. Find the equation of the normal to $y^2=8x$ parallel to the line $y=4x+1$.
6. Find the condition that the line $Ax+By=1$ may be normal to the curve $y=x^3$.

4.7. ROLLE'S* THEOREM AND LAGRANGE'S MEAN VALUE THEOREM

We shall now discuss a simple, yet very important theorem, known as Rolle's Theorem, which gives a set of sufficient conditions

---

*MICHAEL ROLLE (1652-1719) was a member of the French Academy. He contributed mostly to analytic geometry and calculus.

under which the graph of a real function may have a tangent parallel to the x-axis. A simple consequence of Rolle's theorem is Lagrange's mean value theorem which has important applications some of which we shall discuss in the next section.

Theorem 4·1. (*Rolle's Theorem*). *Let f be a function defined on $[a, b]$ such that*

(*i*) *f is continuous on $[a, b]$;*

(*ii*) *f is derivable on $]a, b[$;*

(*iii*) *$f(a)=f(b)$*

Then there exists a real number c between a and b such that $f'(c)=0$.

We shall not give a proof of the above theorem but will confine ourselves only to its consequences.

Remark. Rolle's theorem ensures the existence of *at least one* real number c such that $f'(c)=0$. It does not say anything about the existence or otherwise of more than one such number. As we shall see in Example 20, for a given f, there may exist several numbers c such that $f'(c)=0$.

Example 16. *Verify the hypothesis and conclusion of Rolle's theorem for the function f defined in $[-\pi/2, \pi/2]$ by setting $f(x)=\cos x$.*

Solution. Here we have $f(x)=\cos x$.

(*i*) The function f is continuous in $[-\pi/2, \pi/2]$.

(*ii*) $f'(x)$ exists in $]-\pi/2, \pi/2[$.

(*iii*) $f(-\pi/2)=0=f(\pi/2)$.

Therefore all the conditions in the hypotheses of Rolle's theorem are satisfied.

According to the conclusion of Rolle's theorem, we must have a 'c' in $]-\pi/2, \pi/2[$ such that $f'(c)=0$. This is indeed the case, for $f'(x)=-\sin x$ which vanishes at $x=0$. Observe that here $c=0$ which lies in $]-\pi/2, \pi/2[$.

Example 17. *Examine the validity of the hypotheses and the conclusion of Rolle's theorem in $[-1, 1]$ for the function f defined by $f(x)=|x|$.*

Solution. Observe that

(*a*) f is continuous in $[-1, 1]$.

(*b*) $f(-1)=|-1|=1$.
$f(1)=|1|=1$,
so that $f(-1)=f(1)$.

(*c*) f is not derivable in $]-1, 1[$ in as much as $f'(x)$ does not exist at $x=0$.

Thus all the hypotheses of Rolle's theorem are not satisfied. While two of them are satisfied, one of them is violated.

$$\text{Since } f(x)=\begin{cases} x & \text{if } x \geqslant 0 \\ -x & \text{if } x<0, \end{cases}$$

therefore $f'(x)=-1$, if $x<0$ and $f'(x)=1$ if $x>0$. We find that $f'(x)$ does not vanish anywhere on **R**, and therefore, in particular, it does not vanish anywhere in $]-1, 1[$.

Thus the conclusion of Rolle's theorem is not valid.

Example 18. *Prove that if $a_0, a_1, \ldots, a_n$ be real numbers such that*

$$\frac{a_0}{n+1}+\frac{a_1}{n}+\ldots\ldots+\frac{a_{n-1}}{2}+a_n=0$$

then there exists at least one real number x between 0 and 1 such that

$$a_0x^n+a_1x^{n-1}+\ldots\ldots a_n=0.$$

Solution. Consider the function f defined on $[0, 1]$ by setting

$$f(x)=a_0\frac{x^{n+1}}{n+1}+a_1\frac{x^n}{n}+\ldots\ldots+a_nx,$$

f satisfies all the hypotheses of Rolle's theorem. For

(*i*) f is a polynomial in x, and therefore it is continuous in $[0, 1]$;

(*ii*) Since every polynomial is derivable, therefore $f'(x)$ exists in $]0, 1[$.

(*iii*) $f(0)=0$.

$$\text{Also } f(1)=\frac{a_0}{n+1}+\frac{a_1}{n}+\ldots\ldots+\frac{a_{n-1}}{2}+a_n,$$

$$=0 \text{ (given)},$$

so that $f(0)=f(1)$.

By Rolle's theorem $f'(x)$ must vanish atleast once in $[0, 1]$.

Since $f'(x)=a_0x^n+a_1x^{n-1}+\ldots\ldots+a_n$,

it follows that we must have

$$a_0x^n+a_1x^{n-1}+\ldots\ldots+a_n=0$$

for some real number lying between 0 and 1.

Example 19. *Verify that the function f defined by $f(x)=\sin x$ satisfies all the hypotheses of Rolle's theorem in $[0, 2\pi]$ and show that $f'(x)$ vanishes twice in $]0, 2\pi[$.*

Solution. f satisfies all the hypotheses of Rolle's theorem in $[0, 2\pi]$ because

(*i*) $\sin x$ is continuous in $[0, 2\pi]$,

(*ii*) $\sin x$ is derivable in $]0, 2\pi[$,

(*iii*) $\sin x$ has the value 0 at $x=0$ and $x=2\pi$. Also, $f'(x)=\cos x$ which vanishes twice in $]0, 2\pi[$, namely when $x=\pi/2$ and $x=3\pi/2$.

Geometrical significance of Rolle's theorem

Interpreted geometrically, Rolle's theorem says that the curve

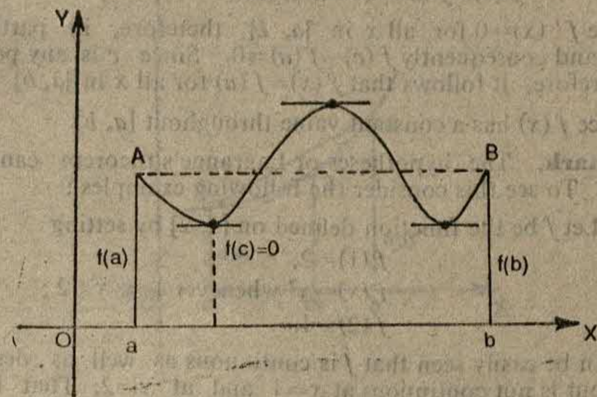


Fig. 4·6.

representing the graph of the function f must have a tangent parallel to x-axis, at least at one point between a and b.

Lagrange's mean value theorem

One important consequence of Rolle's theorem is the following theorem due to Lagrange (1736-1813) which is called Lagrange's theorem.

We shall state the theorem without proof and discuss its implications :

Theorem 4·2. (*Lagrange's mean value theorem*). *Let f be a function defined on $[a, b]$, such that*

(*i*) *f is continuous on $[a, b]$, and*

(*ii*) *f is derivable on $]a, b[$.*

Then there exists a real number c in $]a, b[$ such that

$$f(b)-f(a)=(b-a)f'(c).$$

From Lagrange's theorem we can deduce the following important corollary.

Corollary. *If f is defined and continuous on $[a, b]$ and is derivable on $]a, b[$, and if $f'(x)=0$ for all x in $]a, b[$, then $f(x)$ has a constant value throughout $[a, b]$.*

Proof. Let c be any point of $]a, b]$. Then

(*i*) f is continuous on $[a, c]$;

(*ii*) f is derivable on $]a, c[$.

Since f satisfies all the conditions of Lagrange's mean value theorem on $[a, c]$. Therefore, there exists a real number d between a and c such that

$$f(c)-f(a)=(c-a)f'(d).$$

Since $f'(x)=0$ for all x in $]a, b[$, therefore, in particular, $f'(d)=0$ and consequently $f(c)-f(a)=0$. Since c is any point of $]a, b]$, therefore, it follows that $f(x)=f(a)$ for all x in $]a, b]$.

Hence $f(x)$ has a constant value throughout $[a, b]$.

Remark. The hypotheses of Lagrange's theorem cannot be weakened. To see this consider the following examples :

1. Let f be the function defined on $[1, 2]$ by setting

$$f(1)=2,$$
$$f(x)=x^2 \text{ whenever } 1<x<2\ ;$$
$$f(2)=1.$$

It can be easily seen that f is continuous as well as derivable on $]1, 2[$ but is not continuous at $x=1$ and at $x=2$. That is, the first of the two conditions is violated.

Also, $$\frac{f(2)-f(1)}{2-2}=-1,$$

$$f'(x)=2x, \text{ whenever } 1>x<2,$$

so that $f'(x)$ is positive for all x in $]1, 2[$.

Thus $\dfrac{f(2)-f(1)}{2-1}\neq f(x)$ for any x in $]1, 2[$.

2. Let f be the function defined on $[-1, 2]$ by setting

$$f(x)=|x|, \text{ for all } x \text{ in } [-1, 2].$$

Here f is continuous on $[-1, 2]$, and derivable at all points of $]-1, 2[$ except at $x=0$ (so that the second of the two conditions is violated).

Now $$f'(x)=\begin{cases}-1, \text{ if } x\in]\ 1, 0\ , \\ 1, \text{ if } x\in]0, 2[.\end{cases}$$

Also, $$\frac{f(2)-f(-1)}{2-(-1)}=\frac{1}{3},$$

so that $\dfrac{f(2)-f(-1)}{2-(-1)}\neq f'(x)$ for any x in $]-1, 2[$.

Geometrical Interpretation of Lagrange's theorem

Interpreted geometrically, Lagrange's mean value theorem says (Fig. 4·7) that the tangent to the graph of f at some suitable point

between a and b is parallel to the chord joining the points on the graph with abscissae a and b.

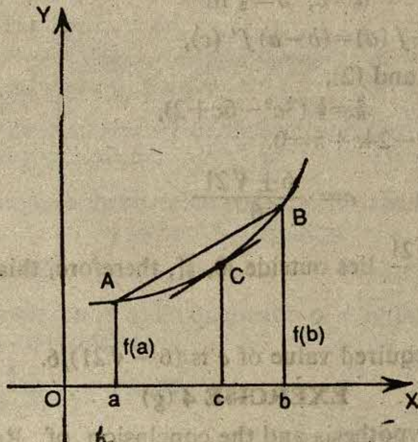


Fig. 4·7. (*a*)

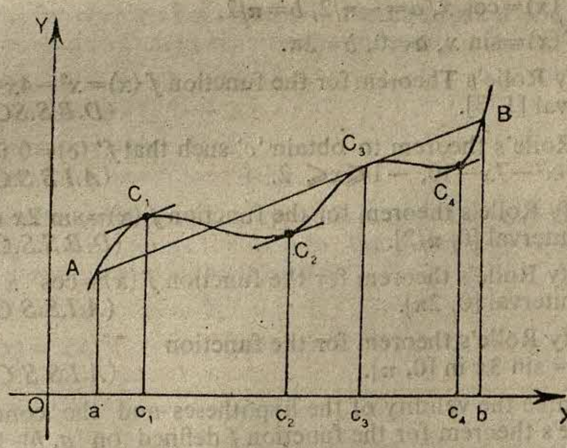


Fig. 4·7. (*b*)

Example 20. *Find 'c' of Lagrange's mean value theorem if* $f(x)=x(x-1)(x-2)$; $a=0$, $b=\frac{1}{2}$.

Solution. $f(x)=x(x-1)(x-2)$.

$\therefore\ f(0)=0, f(\frac{1}{2})=\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)=\frac{3}{8}$. ...(1)

Also, $f'(x)=3x^2-6x+2$.

Putting $a=0,\ b=\frac{1}{2}$ in

$$f(b)-f(a)=(b-a)f'(c),$$

we have form (1) and (2),

$$\frac{3}{8}=\frac{1}{2}(3c^2-6c+2),$$

or $$12c^2-24c+5=0.$$

$$\therefore\qquad c=\frac{6\pm\sqrt{21}}{6}$$

Since $\frac{6+\sqrt{21}}{6}$ lies outside $]0, \frac{1}{2}[$, therefore, this value of c has to be discarded.

Hence the required value of c is $(6-\sqrt{21})/6$.

EXERCISE 4 (*g*)

1. Verify the hypothesis and the conclusion of Rolle's theorem for the function f defined on $[a, b]$ in each of the following cases :
   (*a*) $f(x)=x^2, a=-1, b=1$.
   (*b*) $f(x)=x^3-6x^2+11x-6, a=1, b=3$.
   (*c*) $f(x)=\cos x, a=-\pi/2, b=\pi/2$.
   (*d*) $f(x)=\sin x, a=0, b=2\pi$.
2. Verify Rolle's Theorem for the function $f(x)=x^2-4x+3$ in the interval [1, 3]. (*D.B.S.SC.E. 1985*)
3. Use Rolle's theorem to obtain 'c' such that $f'(c)=0$ for $f(x)=x^3+4x^2-7x-10, -1\leqslant x\leqslant 2$. (*A.I.S.S.C.E. 1986*)
4. Verify Rolle's theorem for the function $f(x)=\sin 2x$ defined in the interval $[0, \pi/2]$. (*D.B.S.S.C.E. 1989*)
5. Verify Rolle's theorem for the function $f(x)=\cos x+\sin x$ in the interval $[0, 2\pi]$. (*A.I.S.S.C.E. 1988*)
6. Verify Rolle's theorem for the function $f(x)=\sin 3x$ in $[0, \pi]$. (*A.I.S.S.C.E. 1987*)
7. Examine the validity of the hypotheses and the conclusion of Rolle's theorem for the function f defined on $[a, b]$ in each of the following cases :
   (*a*) $f(x)=|x|, a=-1, b=1$.
   (*b*) $f(x)=2+(x-1)^{2/3}, a=0, b=2$.
   (*c*) $f(x)=x^{5/2}, a=1, b=5$.
   (*d*) $f(x)=(x-2)\sqrt{x}, a=0, b=2$.

8. Verify the hypotheses and conclusion of Lagrange's mean value theorem for the function f defined on $[a, b]$ in each of the following cases :

(a) $f(x)=x^3$, $a=-2$, $b=1$.

(b) $f(x)=1/x$, $a=1$, $b=4$.

(c) $f(x)=x^n$ (n being a positive integer), $a=-1$, $b=1$.

(d) $f(x)=\cos x$, $a=0$, $b=\pi/2$.

9. Verify Lagrange's mean value theorem for the function

$$f(x)=2x^2-3x+1, \; x\in[1, 3]. \qquad \textit{(D.B.S.S.C.E. 1987)}$$

10. Verify Lagrange's mean value theorem for the function

$$f(x)=2x^2-10x+29$$

in the interval [2, 7]. *(A.I.S.S.C.E. 1984)*

11. Verify the hypothesis and conclusion of Lagrange's mean value theorem for the function

$$f(x)=\frac{1}{4x-1}, \; 1\leqslant x\leqslant 4.$$

(A.I.S.S.C.E. 1989)

12. Use Lagrange's mean value theorem to determine a point P on the curve $y=\sqrt{(x^2-4)}$ in the interval [2, 4] where the tangent is parallel to the chord joining the end-points of the curve. *(A.I.S.S.C.E. 1986)*

13. Find the number 'c' that appears in the conclusion of Lagrange' mean value theorem in each of the following cases :

(a) $f(x)=x^2-2x+3$, $a=1$, $b=\frac{3}{2}$.

(b) $f(x)=x^2$, $a=1$, $b=1{\cdot}2$.

(c) $f(x)=\ln x$, $a=1$, $b=1{\cdot}1$.

(d) $f(x)=\ln x$, $a=1$, $b=1+1/e$.

14. Examine the validity of the hypotheses and the conclusion of Lagrange's mean value theorem for the function f defined on $[a, b]$ in each of the following cases :

(a) $f(x)=|x|$, $a=-2$, $b=1$.

(b) $f(x)=1/x$, if $x\neq 0$, $f(0)=0$, $a=-1$, $b=2$.

(c) $f(x)=x^{1/3}$, $a=-1$, $b=1$.

(d) $f(x)=1+x^{2/3}$, $a=-8$, $b=1$.

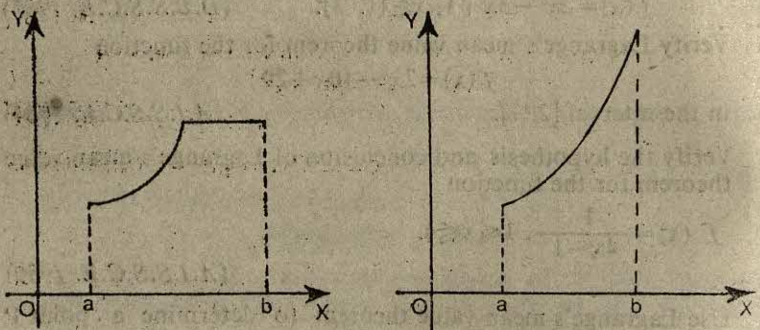
4·8. INCREASING AND DECREASING FUNCTIONS AND SIGN OF THE DERIVATIVE

It is often possible to obtain valuable information about a function from a knowledge of the sign of the derivative. In the present section we shall see how the mean value theorem proves useful in this context. We shall first recall a few definitions.

Definition 4·3. *A fuuction f defined on an interval I is said to be increasing in I iff $f(x_1) \leqslant f(x_2)$ whenever x_1, x_2 are in I and $x_1 < x_2$.*

Definition 4·4. *A function f defined in an interval I is said to be strictly increasing in I if $f(x_1) < f(x_2)$ whenever x_1, x_2 are in I and $x_1 < x_2$.*

Figures 4.8(*a*) and 4.8(*b*) show graphs of an increasing function and a strictly increasing function respectively.



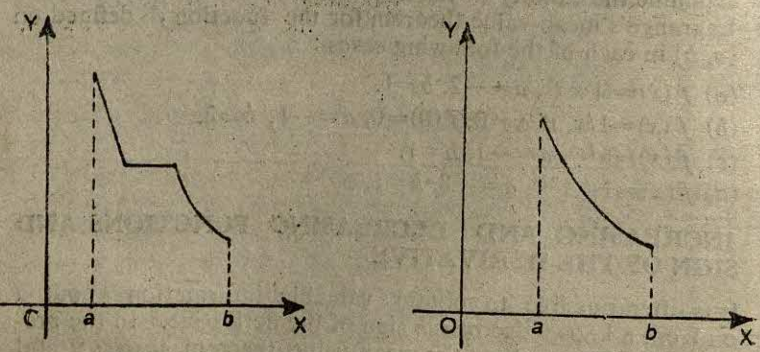
(*a*) (*b*)

Fig. 4·8.

Definition 4·5. *A function f defined on an interval I is said to be decreasing in I if $f(x_1) \geqslant f(x_2)$ whenever x_1, x_2 are in I and $x_1 < x_2$.*

Definition 4·6. *A function f defined on an interval I is said to be strictly decreasing in I if $f(x_1) > f(x_2)$ whenever x_1, x_2 are in I and $x_1 < x_2$.*

Figures 4.9(*a*) and 4.9(*b*) show graphs of a decreasing function and a strictly decreasing function respectively.



(*a*) (*b*)

Fig. 4.9.

Definition 4.7. *A function is said to be monotone in an interval I if it is either increasing in I or decreasing in I.*

Definition 4.8. *A function is said to be strictly monotone in an interval I if it is either strictly increasing in I or strictly decreasing in I.*

Figures 4.8(*a*), 4.8(*b*), 4.9(*a*) and 4.9(*b*) show graphs of functions which are monotone in $[a, b]$. Figures 4.8(*b*) and 4.9(*b*) show graphs of functions which are strictly monotone in $[a, b]$.

Theorem 4·3. *If f is continuous on* $[a, b]$ *and* $f'(x)\leqslant 0$ *in* $]a, b[$, *then f is increasing in* $[a, b]$.

Proof. Let x_1 and x_2 be any two distinct points of $[a, b]$ such that $x_1 < x_2$. Then f satisfies the hypothesis of the mean value theorem in $[x_1, x_2]$. Therefore, there exists a number c such that $x_1 < c < x_2$, and

$$f(x_2)-f(x_1)=(x_2-x_1)f'(c).$$

Since $x_2-x_1>0$ and $f'(c)\geqslant 0$ (for $f'(x)\geqslant 0$ whenever x is in $]a, b[$ and c is a point of $]a, b[$), therefore,

$$f(x_2)-f(x_1)\geqslant 0,$$

i.e., $$f(x_1)\leqslant f(x_2).$$

Since $x_1 < x_2 \Rightarrow f(x_1)\leqslant f(x_2)$ for all x_1, x_2 in $[a, b]$, therefore it follows by definition 4.3 that f is increasing in $[a, b]$.

Theorem 4·4. *If f is continuous on* $[a, b]$ *and* $f'(x)>0$ *in* $]a, b[$, *then f is strictly increasing in* $[a, b]$.

Proof. Let x_1 and x_2 be any two distinct points of $[a, b]$ such that $x_1 < x_2$. Then f satisfies the hypotheses of the mean value theorem in $[x_1, x_2]$. Therefore, there exists a uumber c such that $x_1 < c < x_2$,

and $$f(x_2)-f(x_1)=(x_2-x_1)f'(c).$$

Since $x_2-x_1>0$ and $f'(c)>0$ (why ?),

therefore, $$f(x_2)-f(x_1)>0.$$

Since $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$ for all x_1 and x_2 in $[a, b]$, therefore f is strictly increasing in $[a, b]$.

Theorem 4·5. *If f is continuous on* $[a, b]$ *and* $f'(x)\leqslant 0$ *in* $]a, b[$, *then f is decreasing in* $[a, b]$.

Proof. Let us define a function g by setting

$$g(x)=-f(x) \text{ for all } x \text{ in } [a, b].$$

Then g is continuous on $[a, b]$, and

$$g'(x)=-f'(x)\geqslant 0 \text{ in }]a, b[.$$

Therefore, by Theorem 4·3, the function g is increasing in $[a, b]$. This means that if x_1 and x_2 be any two distinct points of $[a, b]$ such that $x_1 < x_2$, then

$$g(x_1) \leqslant g(x_2),$$

i.e., $\quad -f(x_1) \leqslant -f(x_2),$

i.e., $\quad f(x_1) \geqslant f(x_2).$

Since $x_1 < x_2 \Rightarrow f(x_1) \geqslant f(x_2)$ for all x_1 and x_2 in $[a, b]$, therefore, f is decreasing in $[a, b]$.

Theorem 4·6. *If f is continuous on $[a, b]$ and $f'(x) < 0$ in $]a, b[$, then f is strictly decreasing in $[a, b]$.*

Proof. Let us define a function g by setting

$$g(x) = -f(x) \text{ for all } x \text{ in } [a, b].$$

Then g is continuous on $[a, b]$ and $g'(x) = -f'(x) > 0$ for all x in $]a, b[$. Therefore, by Theorem 4·4, the function g is strictly increasing in $[a,b]$. This means that if x_1 and x_2 be any two distinct points of $[a, b]$ such that $x_1 < x_2$, then

$$g(x_1) < g(x_2),$$

i.e., $\quad -f(x_1) < -f(x_2),$

i.e., $\quad f(x_1) > f(x_2).$

Since $x_1 < x_2 \Rightarrow f(x_1) > f(x_2)$ for all x_1 and x_2 in $[a, b]$, therefore f is strictly decreasing in $[a, b]$.

Remark. We have deduced theorems 4·5 and 4·6 from theorems 4·3 and 4·4 respectively. We could as well have imitated the proofs of theorems 4·3 and 4·4 to prove theorems 4·5 and 4·6.

Having read theorems 4·3 to 4·6, the reader might be wondering as to what would happen if the derivative of a function vanishes over an interval. The situation is covered by the following (corollary to theorem 4·2, page 240) which we state here for the sake of completeness :

If f is continuous on $[a, b]$ and derivable on $]a, b[$, and if $f'(x) = 0$ for all x in $]a, b[$, then f has a constant value throughout $[a, b]$.

The following table summarizes the results of the above theorems for a function f which is given to be continuous on $[a, b]$.

Nature of $f'(x)$ in $]a, b[$	*Nature of f in $[a, b]$*
$f'(x) \geqslant 0$	f is increasing
$f'(x) > 0$	f is strictly increasing
$f'(x) \leqslant 0$	f is decreasing
$f'(x) < 0$	f is strictly decreasing
$f'(x) = 0$	f is constant

Example 21. *Show that the function f, defined on* $\mathbf{R}$ *by*

$$f(x) = x^3 - 6x^2 + 12x - 4 \text{ for all } x \in \mathbf{R}.$$

is increasing in every interval.

Solution. $f(x)=x^3-6x^2+12x-4$,

$$f'(x)=3x^2-12x+12$$
$$=3(x-2)^2.$$

Thus $f'(x)=0$ when $x=2$, and $f'(x)>0$ when $x\neq 2$.

Let c be any real number less than 2. Then f is continuous in $[c, 2]$ and $f'(x)>0$ in $]c, 2[$. Consequently f is increasing in $[c, 2]$.

Similarly f is increasing in every interval $[2, d]$, where d is any real number greater than 2.

From the above results we find that f is increasing in every interval.

Example 22. *Separate the intervals in which the function f defined on* $\mathbf{R}$ *by*

$$f(x)=x^3-6x^2+9x+1 \text{ for all } x\in\mathbf{R},$$

is increasing or decreasing.

Solution. $f(x)=x^3-6x^2+9x+1$.

$\therefore \quad f'(x)=3x^2-12x+9,$

$$=3(x-1)(x-3),$$

so that $\quad f'(x)>0$ whenever $x<1$,

$f'(x)=0$ when $x=1$,

$f'(x)<0$ whenever $1<x<3$,

$f'(x)=0$ whenever $x=3$,

$f'(x)>0$ whenever $x>3$.

Since $f'(x)>0$ in each of the intervals $]-\infty, 1[$ and $]3, \infty[$, and since f is continuous everywhere, therefore it is strictly increasing in each of the intervals $]-\infty, 1]$ and $[3, \infty[$.

Also, since $f'(x)<0$ in $]1, 3[$ and since f is continuous in $[1, 3]$, therefore it is strictly decreasing in $[1, 3]$.

EXERCISE 4(h)

1. Show that the function f, defined on $\mathbf{R}$ by
$$f(x)=x^3+3x^2+3x-8 \text{ for all } x\in\mathbf{R}$$
is increasing in every interval.

2. Show that the function f, defined on $\mathbf{R}$ by
$$f(x)=9-12x+6x^2-x^3 \text{ for all } x\in\mathbf{R}$$
is decreasing in every interval.

3. Separate the intervals in which the function f, defined on $\mathbf{R}$ by
$$f(x)=2x^3-15x^2+36x-7 \text{ for all } x\in\mathbf{R}$$
is increasing or decreasing.

Determine the intervals in the function f defined below is increasing or decreasing :

4. $f(x)=2x^3-24x+107$. (*A.I.S.S.C.E., 1984*)
5. $f(x)=x^3+5x^2-1$. (*D.B.S.S.C.E., 1984*)
6. $f(x)=x^4-8x^3+22x^2-24x+21$. (*A.I.S.S.C.E., 1984*)
7. $f(x)=\cos x,\ 0\leqslant x\leqslant 2\pi$. (*D.B.S.S.C.E., 1985*)
8. $f(x)=2x^2-24x+5$. (*A.I.S.S.C.E., 1985*)
9. $f(x)=2x^3-3x^2-12x+6$. (*D.B.S S.C.E., 1986*)
10. Show that the function
$$f(x)=x^3-6x^2+15x+3$$
is an increasing function for all x. (*A.I.S.S.C.E., 1988*)
11. Determine, for what values of x, the function
$$f(x)=x^4-\frac{x^3}{3}$$
is increasing or decreasing. At what point is the tangent parallel to the x-axis. (*D.B.S.S.C.E., 1986*)
12. Determine, for which values of x, the function
$f(x)=2x^3-15x^2+36x+1$ is increasing
and for which values it is decreasing.

Also determine the points where the tangents to the graph of the function are parallel to the x-axis. (*A.I.S.S.C.E., 1986*)
13. Determine the intervals in which the function
$$x^3-9x^2+15x+11$$
is increasing or decreasing. Also find the points on the graph at which the tangents are parallel to the x-axis.
(*D.B.S.S.C.E., 1987*)
14. Determine, for which values of x, the function
$$f(x)=\frac{x}{x^2+1}$$
is increasing, and for which values of x, it is decreasing. Find also the points at which the tangent is parallel to the x-axis.
(*A.I.S.S.C.E., 1987*)
15. Determine the intervals in which the function
$$f(x)=\frac{x^3}{3}+\frac{x^2}{2}-2x+11$$
is increasing or decreasing. At what points are the tangents to the graph of the function parallel to the x-axis ?
(*D.B S.S.C.E., 1988*)

4·9. MAXIMA AND MINIMA

Consider the following functions :

(*a*) $f(x)=x^3$, for all $x \in \mathbf{R}$

(*b*) $g(x)=x^4$, for all $x \in \mathbf{R}$

(*c*) $h(x)=1/(1+x^2)$, for all $x \in \mathbf{R}$

(*d*) $s(x)=\sin x$, for all $x \in \mathbf{R}$

What can you say about the geeatest and least values of these functions ? Let us consider these functions one by one :

(*a*) f is strictly increasing on $\mathbf{R}$. It can take every real number as a value for some value of x. In other words, it has neither a maximum value nor a minimum value.

(*b*) $g(x) \geqslant 0$ for all $x \in \mathbf{R}$ and $g(0)=0$. Thus the minimum value of g is 0. It is attained when $x=0$. Also, since $g(x)$ can be made as large as we please by taking x sufficiently large, therefore g does not attain any maximum.

(*c*) $h(x) \leqslant 1$ for all $x \in \mathbf{R}$ and $h(0)=1$. Therefore the maximum value of h is 1. It is attalned when $x=1$. Also, since $h(x)>0$ for all $x \in \mathbf{R}$ and can be made as close to 0 as we please by taking h sufficiently large, therefore h does not attain any minimum.

(*d*) Since $-1 \leqslant s(x) \leqslant 1$, and $s(-\pi/2)=-1$, $s(\pi/2)=1$, therefore s attains both a minimum as well as a maximum.

Let us make the meaning of maximum and minimum values as used in the above illustrations precise by making the following definitions :

Definition 4·9. *A function f defined on an interval I is said to have a maximum value at $x=c$ if $c \in I$ and $f(x) \leqslant f(c)$ for all $x \in I$.*

Definition 4·10. *A function f defined on an interval I is said to have a minimum value at $x=c$ if $c \in I$ and $f(x) \geqslant f(c)$ for all $x \in I$.*

Remark. The maximum value as described in def. 4·9 above is sometimes called an *absolute maximum* or a *global maximum* to distinguish it from *local maximum* to be introduced a little later in this section. Similarly the minimum value as described in def. 4·10 above is sometimes called an *absolute minimum* or a *global minimum* to distinguish it from *local minimum* to be introduced a little later in this section.

Example 23. *Find the maximum and minimum values, in case they exist, for each of the following functions :*

(*a*) $f(x)=2+2x-x^2$ *for all* $x \in \mathbf{R}$

(*b*) $g(x)=x^2-6x-17$ *for all* $x \in \mathbf{R}$.

Solution. (*a*) Writing $2+2x-x^2=3-(x-1)^2$, we find that $f(x) < 3$ for all $x \in \mathbf{R}$ and $f(1)=3$. Since $f(x) \leqslant f(1)$ for all $x \in \mathbf{R}$, therefore f has (an absolute) maximum at $x=1$, the maximum value being 3. Also, since $f(x)$ can be made less than any number whatever by taking x sufficiently large, therefore f does not have any minimum.

(*b*) Writing $x^2-6x-17=(x-3)^2-26$, we find that $g(x) \geqslant -26$ for all $x \in \mathbf{R}$ and $g(3)=-26$. Since $g(x) \geqslant g(3)$ for all $x \in \mathbf{R}$, therefore g has (an absolute) minimum at $x=3$, the minimum value being -26. Also, since $g(x)$ can be made as large as we please by taking x sufficiently large, therefore g does not have any maximum.

In the above example we were able to examine the maxima/minima of functions rather easily. This may not always be so. In case we cannot decide about the maxima/minima of a function over an interval I by inspection, we proceed in a systematic way as follows :

Step 1. Separate the interval I over which we have to find the maximum (resp. minimum) value of a function f into intervals over which the function is increasing/decreasing.

Step 2. (*a*) If f is increasing over an interval $]-\infty, a]$ it does not have a minimum in this interval but has a maximum at $x=a$.

(*b*) If f is decreasing in $]-\infty, a]$ it does not have a maximum in this interval but has a minimum at $x=a$.

(*c*) If f is increasing in $[a, b]$, then it has a minimum at $x=a$ and maximum at $x=b$.

If f is decreasing in $[a, b]$, then it has a maximum at $x=a$ and minimum at $x=b$.

(*d*) If f is increasing in $[b, \infty[$, then it has a minimum at $x=b$ but no maximum.

If f is decreasing in $[b, \infty[$, then it has a maximum at $x=b$ but no mininum.

Step 3. The absolute maximum (resp. minimum) of f over I is the largest (resp. smallest) of all the maxima (resp. minima) of f over the various sub-intervals obtained in step 1. If f does not have an absolute maximum (resp. minimum) over any one of the sub-intervals, then it does not have an absolute maximum (resp. minimum) over I.

Example 24. *Find the absolute maximum and minimum of the function f defined by*

$$f(x)=2x^3-9x^2+12x+20$$

(*a*) *over* $\mathbf{R}$ (*b*) *in the interval* $[-3, 3]$.

Solution. **Step 1.** $f(x)=2x^3-9x^2+12x+20$

$$f'(x)=6x^2-18x+12=6(x-1)(x-2)$$

so that

$$\begin{aligned} f'(x)&>0 && \text{whenever } x<1, \\ f'(x)&=0 && \text{when } x=1 \\ f'(x)&<0 && \text{whenever } 1<x<2 \\ f'(x)&=0 && \text{whenever } x=2 \\ f'(x)&>0 && \text{whenever } x>2 \end{aligned}$$

Since $f'(x)>0$ in each of the intervals $[-\infty, 1[$ and $]2, \infty[$ and sidce f is continuous everywhere, therefore f is strictly increasing in each of the intervals $]-\infty, 1]$ and $[2, \infty]$.

Also, since $f'(x)<0$ in $]1, 2[$ and since f is continuous in $[1, 2]$, therefore f is strictly decreasing in $[1, 2]$.

Step 2. Since f is strictly increasing in $]-\infty, 1]$, therefore it does not have a minimum in this interval.

Also, since f is strictly increasing in $[2, \infty[$, therefore it does not have a maximum in this interval.

Step 3. f does not have an absolute maximum or an absolute minimum over **R**.

(*b*) **Step 1.** Using the calculations in step 1 of (*a*) above and considering only the interval $[-3, 3]$, we find that f is strictly increasing in each of the interval $[-3, 1]$, and $[2, 3]$, and is strictly decreasing in the interval $[1, 2]$.

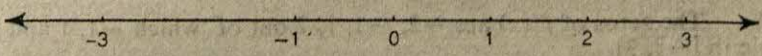


Fig. 4·10.

Step 2. Since f is strictly increasing in $[-3, 1]$, therefore

$$\left.\begin{aligned} &\text{the minimum value of } f \text{ in } [-3, 1]=f(-3), \\ \text{and } &\text{the maximum value of } f \text{ in } [-3, 1]=f(1), \end{aligned}\right\} \qquad \text{...(A)}$$

Since f is strictly decreasing in $[1, 2]$, therefore

$$\left.\begin{aligned} &\text{the minimum value of } f \text{ in } [1, 2]=f(2), \\ \text{and } &\text{the maximum value of } f \text{ in } [1, 2]=f(1). \end{aligned}\right\} \qquad \text{...(B)}$$

Again, since f is strictly increasing in $[2, 3]$, therefore

$$\left.\begin{aligned} &\text{the minimum value of } f \text{ in } [2, 3]=f(2) \\ &\text{the maximum value of } f \text{ in } [2, 3]=f(3) \end{aligned}\right\} \qquad \text{...(C)}$$

Step 3. From (A), (B) and (C) above, we find that the minimum value of f is the *smallest* of the numbers $f(-3)$ and $f(2)$.

Since $f(-3)=2(-3)^3-9(-3)^2+12(-3)+20=-151$,

$f(2)=2(2)^3-9(2)^2+12(2)+20=24$,

it follows that f has an absolute minimum at $x=-3$ and the minimum value is -151.

Again, from (A), (B) and (C) above, we find that the maximum value of f is the biggest of the numbers $f(1)$ and $f(3)$.

Since $f(1)=2(3)^3-9(1)^2+12(1)+20=25$,

$$f(3)=2(3)^3-9(3)^2+12(3)+20=29,$$

therefore f has an absolute maximum at $x=3$ and the maximum value is 29.

Remark. In view of the above discussion we have the following rule for finding the absolute minimum/maximum of a differentiable function f over a closed and bounded interval $[a, b]$:

If f is differentiable on $[a, b]$, and $c_1, c_2, \dots, c_n$ are the zeros of f' lying in $[a, b]$, then the absolute minimum of f over $[a, b]$ is

$$\text{min. } \{f(a), f(c_1), \dots, f(c_n), f(b)\},$$

and the absolute maximum of f over $[a, b]$ is

$$\text{max. } \{f(a), f(c_1), \dots, f(c_n), f(b)\}.$$

Example 25. *Find the absolute minimum and maximum of the function f defined by*

$$f(x)=3x^5-25x^3+60x+15$$

in the interval $[-\frac{3}{2}, 3]$.

Solution. Since f is a polynomial function, therefore it is differentiable on every interval, and in particular over $[-\frac{3}{2}, 3]$.

$$f'(x)=15x^4-75x^2+60,$$
$$=15(x+2)(x+1)(x-1)(x-2).$$

The zeros of $f'(x)$ are $-2, -1, 1, 2$ out of which -1, 1 and 2 lie in $[-\frac{3}{2}, 3]$.

Now $f(-\frac{3}{2})=3(-\frac{3}{2})^5-25(-\frac{3}{2})^3+60(-\frac{3}{2})+15=-429/32$

$$f(-1)=3(-1)^5-25(-1)^3+60(-1)+15=-23$$
$$f(1)=3(1)^5-25(1)^3+60(1)+15=53$$
$$f(2)=3(2)^5-25(2)^3+60(2)+15=31$$
$$f(3)=3(3)^5-25(3)^3+60(3)+15=249$$

Comparing all the above values we find that $f(-1)$ is the smallest and $f(3)$ is the biggest.

Thus f has an absolute minimum at $x=-1$ and absolute maximum at $x=3$. The values at these points are -23 and 249 respectively.

EXERCISE 4 (*i*)

For each of the following functions find the points at which the function has an absolute minimum and absolute maximum in the indicated interval :

1. $f(x)=x^2-7x+6$, $[0, 6]$.
2. $f(x)=3x^2+2x-1$, $[-5, 5]$

3. $f(x)=x^3-3x+1$, $[-4, 4]$.
4. $f(x)=3x^2-4x+1$, $[-1, 1]$.
5. $f(x)=\frac{1}{4}x^4-x^3+\frac{3}{2}x^2-x+1$, $[2, 2]$.
6. Find the absolute maximum and absolute minimum for the function

$$f(x)=3x^5-25x^3+60x+1 \text{ in } [-2, 1].$$

7. Examine the function f defined by $f(x)=2x^3-6x^2+5$ for all $x \in \mathbf{R}$, for absolue maximum and minimum over $\mathbf{R}$.

4·9·1. Local maxima and minima

Consider the function f defined by

$$f(x)+x^3-3x^2+2 \text{ for all } x \in \mathbf{R}.$$

This function does not possess an absolute maximum or absolute minimum over $\mathbf{R}$. However,

$$f'(x)=3x^2-6x=3x(x-2),$$

so that $f'(x)$ vanishes at $x=0$ and $x=2$.

Also $f'(x)>0$ if $x<0$ and $f'(x)<0$ if $0<x<2$. This means that if we take a small interval around the point 0, say $]-\frac{1}{2}, \frac{1}{2}[$, then in this interval $f(x) \leqslant f(2)$, *i.e.*, f has a maximum at $x=0$ so far as this interval is concerned.

Similarly, $f'(x)<0$ if $0<x<2$ and $f'(x)>0$ if $x>2$. This means that if we take a small interval around the point $x=2$, say $]\frac{7}{4}, \frac{9}{4}[$, then in this interval $f(x)>f(2)$, *i.e.*, f has a minimum at $x=2$ so far as this interval is concerned. We say that f has a *local maximum* at $x=0$ and a *local minimum* at $x=2$.

[Note that there is nothing special either about the interval $]-\frac{1}{2}, \frac{1}{2}[$ or about the interval $]\frac{7}{4}, \frac{9}{4}[$.

What we are concerned about are *some* interval around $x=0$ and *some* interval around $x=\frac{1}{2}$ howsoever small these intervals may be. We shall now try to make precise the meaning of local maxima and local minima.

Definition 4·11. *A function f is said to have a local maximum at a point $x=c$, if in some interval I containing $c, f(x)<f(c)$ for all $x \in I$.*

Definition 4·12. *A function f is said to have a local minimum at a point $x=c$, if for some interval I containing $c, f(x)>f(c)$ for all $x \in I$.*

Remark. If f has either a local maximum or a local minimum at $x=c$, then we say that f has an *extreme value* or *extremum* at $x=c$. While a function has at the most only one absolute maximum value and at the most only one absolute minimum value, it may have several local maxima and several local minima.

If a function has a local maximum at a point c, then at points close to c its graph is as shown in Fig. 4·11, $f'(x) > 0$ for values of x close to c but less than it, and $f'(x) < 0$ for values of x close to c but greater than it. This suggests that $f'(c)$ must be zero.

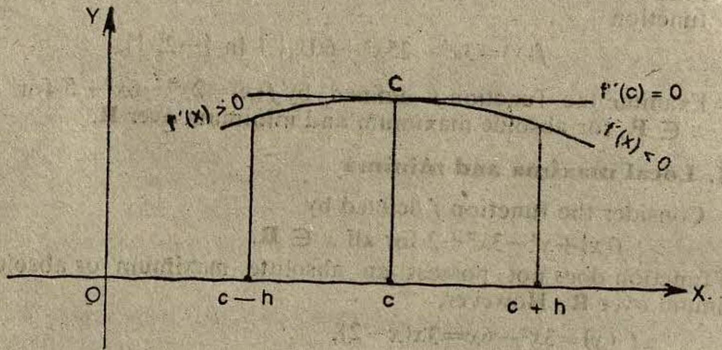


Fig. 4·11.

If a function has a local minimum at a point c, then at points close to c its graph is as shown in Fig. 4·12. $f'(x) < 0$ for values of x close to c but less than it, and $f'(x) > 0$ for values of x close to c but greater than it. This suggests that $f'(c)$ must be zero.

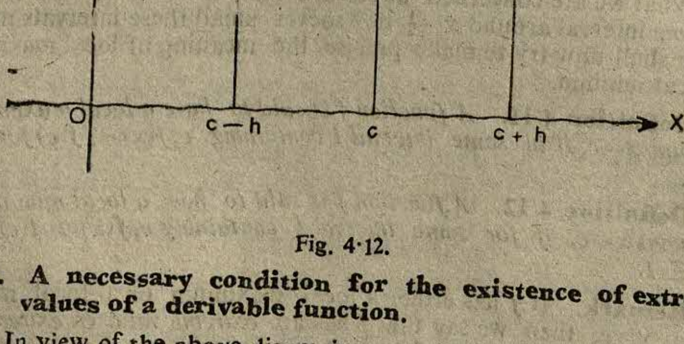


Fig. 4·12.

4·9·2. A necessary condition for the existence of extreme values of a derivable function.

In view of the above discussion, we give below (without proof) a necessary condition for a derivable function to have an extreme value at a point $x=c$.

Theorem 4·7. *Let f be a function defined on an open interval I and let f be derivable at $c \in I$. If f has an extreme value at c, then $f'(c)=0$.*

Remarks 1. The above theorem says that if a function is derivable at a point, then it can have an extreme value at that point only if its derivative at that point vanishes. A function may, however, have an extremum at a point without having a derivative at that point. For example, let f be the function defined on **R** by setting

$f(x)=x^{2/3}$ for all $x \in \mathbf{R}$. f is not derivable at $x=0$, but it has a minimum value at $x=0$.

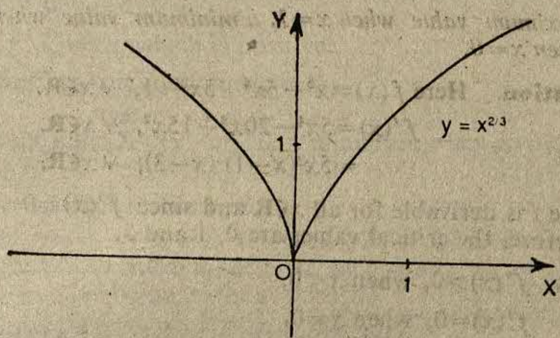


Fig. 4·13.

2. The condition $f'(c)=0$ in Theorem 4·7 for the existence of an extreme value is only necessary. There exist functions for which this condition is satisfied but which do not have an extremum at $x=c$. For example, let f be the function defined on **R** by setting

$f(x)=x^3$ for all $x \in \mathbf{R}$.

Here $f'(0)=0$ but f does not have an extreme value at $x=0$ (see Fig. 4·13).

3. If $f'(x)=0$ at $x=c$, then we say that $x=c$ is a *stationary point* for f. Also, $f'(c)$ is then said to be a *stationary value* of f.

4. In view of the above theorem, we find that if a function f has an exterme value at a point $x=c$, then either f is not derivable at $x=c$, of $f'(c)=0$. Therefore, in order to investigate the maxima and minima of a function f, we have to first find out the values of x for which either $f'(x)$ does not exist, or if $f'(x)$ exists, then it vanishes. (These values are generally called the critical values for the function f.) We then examine as to for which of these values does the function actually have a maximum or a minimum.

4·9·3. Sufficient conditions for the existence of extreme values

In the following theorems we give without proof two sets of sufficient conditions for the existence of extreme values.

Theorem 4·8. (*First derivative test*)

Let f be derivable on an open interval I and let $f'(c)=0$ at some point $c \in I$. If $f'(x)$ changes sign from positive to negative (resp. negative to positive) as x passes through c, then f has a maximum (resp. minimum) at $x=c$.

Example 26. *Show that the function f, defined by*

$$f(x)=x^5-5x^4+5x^3-1 \text{ for all } x \in \mathbf{R},$$

has a maximum value when $x=1$, a minimum value when $x=3$ and neither when $x=0$.

Solution. Here $f(x)=x^5-5x^4+5x^3-1,\ \forall x \in \mathbf{R}.$

$$f'(x)=5x^4-20x^3+15x^2,\ \forall x \in \mathbf{R}.$$
$$=5x^2(x-1)(x-3),\ \forall x \in \mathbf{R}.$$

Since f is derivable for all $x \in \mathbf{R}$ and since $f'(x)=0$ when $x=0$, 1, 3, therefore, the critical values are 0, 1 and 3.

Now $f'(x)>0$, when $x<0$,
$f'(x)=0$, when $x=0$,
$f'(x)>0$, when $0<x<1$,
$f'(x)=0$, when $x=1$,
$f'(x)<0$, when $1<x<3$,
$f'(x)=0$, when $x=3$,
$f'(x)>0$, when $x>3$.

Here $f'(x)$ has the same sign for all x in $]-1, 0[\cup]0, 1[$, and consequently f has neither a maximum value nor a minimum value at $x=0$.

Since $f'(x)$ changes sign from positive to negative as x passes through 1, therefore f has a maximum value at $x=1$.

Again, since $f'(x)$ changes sign from negative to positive as x passes through 3, therefore f has a maximum value at $x=3$.

Example 27. *Given n real numbers $a_1, a_2, \cdots\cdots, a_n$, find the value of x for which the sum*

$$\sum_{i=1}^{n} (x-a_i)^2$$

is a minimum.

Solution. Let us define a function f by setting

$$f(x)=\sum_{i=1}^{n} (x-a_i)^3 \text{ for all } x\in\mathbf{R}.$$

Then $f'(x)=\sum_{i=1}^{n} 2(x-a_i)=2nx-2\sum_{i=1}^{n} a_i=2n(x-\bar{a})$,

where $\bar{a}=(a_1+a_2+...+a_n)/n$ is the arithmetic mean of the numbers $a_1, \cdots, a_n$.

Also, $f''(x)=2n$

Since $f'(x)$ exists for all $x\in\mathbf{R}$, therefore the critical values of x are given by $f'(x)=0$. Since $f'(x)=\Leftrightarrow x=\bar{a}$, therefore, the only critical value of x is $\bar{a}$.

Also $f'(\bar{a})=2n$, which is positive.

Hence f has a minimum at $x=\bar{a}$.

Thus the sum $\sum_{i=1}^{n} (x-a_i)^2$ is a minimum when $x=\bar{a}$.

4·9·4. Second derivative test

Sometimes it is not easy to check whether $f'(x)$ changes sign as x passes through a critical value c. In such situations we use another test, which we give below without proof :

Theorem 4·9. (*Second derivative test*).

Let f be differentiable, on an open-interval I, and let $f'(c)=0$ for some $c\in$I. If (*i*) $f''(c)<0$ then f has a local maximum at $x=c$ (*ii*) $f''(c)>0$, then f has a local minimum at $x=c$.

Remark. It is not possible to draw any conclusion regarding the maximum or minimum values of a function at a point $x=c$ if $f'(c)=0$ and $f''(c)=0$. For example,

(*i*) Let f be the function defined on $\mathbf{R}$ by setting $f(x)=x^3$ for all $x\in\mathbf{R}$. Here $f'(0)=0=f''(0)$, and the function has neither a maximum nor a minimum at $x=0$.

(*ii*) Let f be the function defined on $\mathbf{R}$ by setting $f(x)=x^4$ for all $x\in\mathbf{R}$. Here $f'(0)=0, f''(0)=0$ and the function has a minimum at $x=0$.

(*iii*) Let f be the function defined on $\mathbf{R}$ by setting $f(x)=-x^4$ for all $x\in\mathbf{R}$. Here $f'(0)=0=f''(0)$, and the function has a maximum at $x=0$.

If $f'(c)=0, f''(c)=0$, we may try to apply the first derivative test.

Example 28. *Find the values of x for which the function f, defined by*

$$f(x)=12x^5-45x^4+40x^3+6 \text{ for all } x\in\mathbf{R}$$

has a local maximum or a local minimum.

Solution. Here $f(x)=12x^5-45x^4+40x^3+6$, for all $x\in\mathbf{R}$

$$f'(x)=60x^4-180x^3+120x^2,$$
$$=60x^2(x-1)(x-2).$$
$$f''(x)=240x^3-540x^2+240x,$$
$$=60x(4x^2-9x+4).$$

Since f is derivable for all $x\in\mathbf{R}$, and since $f'(x)=0$ when $x=0, 1, 2$, therefore the critical values are 0, 1, and 2.

Now $f''(0)=0, f''(1)=-60, f''(2)=240$.

Since $f'(1)=0, f''(1)<0$, therefore f has a local maximum at $x=1$.

Again, since $f'(2)=0$, $f''(2)>0$, therefore f has a local minimum at $x=2$.

4·9·5. Application of maxima-minima to problems

We shall now illustrate the application of maxima and minima to solution of problems by means of examples.

Example 29. *A rectangular box with a square base and open at the top is to be made from a piece of card-board having area 192 cm². What should be the dimensions of the box so that the volume of the box may be maximum possible ? (There is to be no wastage and the entire piece of card-board is used.)*

Solution. Let the length of each side of the square base be x cm, and let the height of the box be y cm.

The total surface area of the box

=area of the base+area of the four side-faces

$=x^2+4xy=192$ (given)

$$y=\frac{192-x^2}{4x}=\frac{48}{x}-\frac{x}{4} \qquad ...(1)$$

Also, the volume V of the box, is given by

$$V=x^2y=x^2\left(\frac{48}{x}-\frac{x}{4}\right), \text{ by (1)}$$

$$=48x-\frac{1}{4}x^3. \qquad ...(2)$$

Differentiating both sides of (2), we have

$$\frac{dV}{dx}=48-\frac{3}{4}x^2, \frac{d^2V}{dx^2}=-\frac{3}{2}x. \qquad ...(3)$$

From (3), we find that

$$\frac{dV}{dx}=0 \Rightarrow 48-\frac{3}{4}x^2=0,$$

or $$x^2=64, \text{ so that } x=\pm 8.$$

Now x cannot be negative, therefore $x=8$.

Then $\frac{d^2V}{dx^2}=-\frac{3}{2}\cdot 8=-12<0$. Since $\frac{d^2V}{dx^2}<0$, therefore V is maximum.

From (1), we find that when $x=8$, $y=4$.

Thus, for the volume to be maximum, length of each side of the base must be 8 cm, and height must be 4 cm.

Example 30. *A can in the form of a closed right circular cylinder is to be made of sheet metal, to have capacity V. Find height of the can and the diameter of its base so that the metal used may be a minimum.*

Solution. Let r be the radius of the base, h the height, and S the total surface area of the can.

Then $$V=\pi r^2 h, \qquad \text{...(1)}$$
$$S=2\pi r^2+2\pi rh. \qquad \text{...(2)}$$

Since V is given, we may express h and S as functions of r in the form

$$h=V/(\pi r^2),$$
$$S=2\pi r^2+\frac{2V}{r}. \qquad \text{...(3)}$$

We have to minimize S.

Differentiating (3) w.r.t. r, we have

$$\frac{dS}{dr}=4\pi r-2V/r^2. \qquad \text{...(4)}$$

If S has an extremum, $dS/dr=0$, which yields

$$r=[V/(2\pi)]^{1/3}.$$

Also, $$\frac{d^2S}{dr^2}=4\pi+4V/r^3,$$

which is always positive.

Therefore S is a minimum when $r=[V/(2\pi)]^{1/3}$, and then $h=V/(\pi r^2)=2r$.

Thus the diameter of the base of the can must be $(4V/\pi)^{1/3}$ and its height must be equal to its diameter.

EXERCISE 4 (*j*)

Find the local maxima and minima of each of the following functions :

1. $f(x)=(x-1)(x-2)(x-3)$. (*A.I.S.S.C.E. 1988*)
2. $f(x)=2x^3-21x^2+36x-20$. (*A.I.S.S.C.E. 1984*)
3. $f(x)=-x+2 \sin x,\ 0 \leqslant x < 2\pi$. (*A.I.S.S.C.E. 1984*)
4. $f(x)=\sin 2x-x,\ -\pi/2 < x < \pi/2$. (*D.B.S.S.C.E. 1984*)
5. $f(x)=2 \cos x+x,\ 0 \leqslant x \leqslant \pi$. (*A.I.S.S.C.E. 1987*)
6. $f(x)=\sin x+\frac{1}{2} \cos 2x,\ 0 < x \leqslant \pi/2$.
7. Investigate the maximum and minimum values of the function f, defined by
$$f(x)=x^3-12x^2+45x \text{ for all } x \in [0, 7].$$
8. Determine the values of x for which the function f, defined by
$$f(x)=12x^5-45x^4+40x^3+6, \text{ for all } x \in \mathbf{R}.$$
attains *a* (*i*) maximum value (*ii*) a minimum value.
9. The sum of two numbers is 24. Find the numbers so that their product is maximum. (*D.B.S.S.C.E. 1987*)
10. The product of two numbers is 256. Find the numbers so that their sum is least.
11. The sum of two numbers is 160. If the sum of thrice the square of the first and twice the square of the second is a minimum, find the numbers.
12. Find the dimensions of the rectangle of area 96 cm² whose perimeter is the least. Find also its perimeter. (*D.B.S.S C.E. 1986*)
13. A rectangle has a given perimeter of 40 cm. What should be its dimensions so that its area may be maximum ?
14. A figure consists of a semi-circle with a rectangle on its diameter. Given that the perimeter of the figure is 20 m, find the dimensions of the figure so that its area may be maximum. (*A.I.S.S.C.E. 1986*)
15. The sum of the perimeters of a circle and a square is 300 cm. Show that when the sum of the areas is minimum, each side of the square is double the radius of the circle.
16. A closed right circular cylinder has a surface area 400 cm². What should be its radius and height so as to have the largest possible volume ?
17. A right circular cone is such that the sum of its diameter and height is 18 cm. Find its maximum volume.

18. Show that of all rectangles of given area, the square has the smallest perimeter.

19. Show that of all rectangles having a given perimeter, the square has the largest area. *(D.B.S.S.C.E., 1985)*

20. Show that rectangle of maximum area that can be inscribed in a given circle is a square.

21. Show that rectangle of maximum perimeter which can be inscribed in a circle of radius a is a square of side $a\sqrt{2}$.

22. Prove that the least perimeter of an isosceles triangle in which a circle of radius r can be inscribed is $6\sqrt{3}r$.

23. The sum of the lengths of the hypotenuse and another side of a right angled triangle is given. Prove that the area of the triangle is a maximum when the angle between the hypotenuse and the given side is $\pi/3$.

24. Show that the height of an open right circular cylinder of given total surface and greatest volume is equal to the radius of its base.

25. Show that the radius of the right circular cylinder of greatesr curved surface which can be inscribed in a given right circulat cone is half that of the cone.

26. Show that the altitude of a right circular cone of maximum volume which can be inscribed in a sphere of radius r is $\frac{4}{3}r$.

27. Find the altitude of a right circular cone of maximum curved surface which can be inscribed in a sphere of radius r. *(A.I.S.S.C.E., 1989)*

28. Show that the altitude of a right circular cone of minimum volume circumscribed about a sphere of radius r is $4r$, and its semi-vertical angle is $\sin^{-1}\frac{1}{3}$.

29. Show that the semi-vertical angle of the cone of maximum volume and given slant height is $\tan^{-1}\sqrt{2}$.

30. Given the total surface of a right circular cone, show that when the volume of the cone is maximum then the semi-vertical angle is $\sin^{-1}\frac{1}{3}$.

31. Show that the right circular cone of least curved surface and given volume has an altitude equal to $\sqrt{2}$ times the radius of the base.

4·10. CURVE SKETCHING

In the second chapter we had studied the graphs of some real functions. We had seen how considerations of symmetry, behaviour of the function for small values of x as also for large values of x, and intersection of the graph with the axes of co-ordinates can prove helpful for sketching the graph.

Having learnt some elements of differential calculus in the preceding chapter, we have seen how the concept of the derivative can be used (*i* to separate intervals in which a function is increasing or decreasing, (*ii*) to determine the maxima and minima of a function, and (*iii*) to determine the equations of the tangent and normal to a curve. We shall now try to see as to how we can use this information to sketch a curve which represents the graph of a given function.

While our main emphasis in this section will be on learning the use of derivatives for sketching of curves, for the sake of completeness. We shall enumerate the various points that we should consider while trying to sketch a curve. This will enable us to consolidate the knowledge acquired so far in the matter of curve-sketching.

Sometimes the equation of a curve is given in implicit form as $f(x, y)=0$ instead of the form $y=F(x)$. Since the explicit form $y=F(x)$ can always be put as $f(x, y)\equiv y-F(x)=0$, therefore we shall consider $f(x, y)=0$ in our discussion.

In order to trace a curve whose equation is of the form $f(x, y)=0$, examination of the following facts is often useful :

1. Symmetries. (*a*) A curve $f(x, y)=0$ is said to be symmetrical with respect to the x-axis if $(x, -y)$ is also a point on the graph whenever (x, y) is a point on it, *i.e.*, $f(x, -y)=0\Leftrightarrow f(x, y)=0$, *i.e.*, if its equation remains unaltered when y is replaced by $-y$ throughout. Such a symmetry is often called *reflective symmetry about the x-axis* because the portion of the graph below the x-axis is the reflection in the x-axis of the portion of the graph above the axis.

If a curve is symmetrical with respect to the x-axis, it is enough to first sketch the portion of the graph above the x-axis (i.e., in the first and the second quadrants), and then sketch its reflection in the x-axis.

(*b*) A curve $f(x, y)=0$ is said to be *symmetrical with respect to the y-axis* if $(-x, y)$ is also a point on the graph whenever (x, y) is a point on it, *i.e.* $f(-x, y)=0\Leftrightarrow f(x, y)=0$, *i.e.* if its equation remains unaltered when x is replaced by $-x$ throughout. Such a symmetry is called *reflective symmetry about the y-axis* because the portion of the graph to the left of the x-axis is the reflection in the y-axis of the portion of the graph to the right of the y-axis.

If a curve is symmetrical with respect to the y-axis, it is enough to first sketch the portion of the graph to the right of the y-axis (i.e., in the first and the fourth quadrants), and then sketch its reflection in the y-axis.

(*c*) *If a curve is symmetrical with respect to both the axes, then, it is enough to first sketch the portion of the graph in the first*

quadrant, and then take its reflection in the y-axis to get the portion of the graph in the second quadrant. To complete the graph, we take the reflections in the x-axis of the portions of the graph in the first and the second quadrants.

(d) A curve is said to be symmetrical with respect to the origin if $(-x, -y)$ is also a point on the graph whenever (x, y) is a point on it, *i.e.*, if $f(-x, -y)=0 \Leftrightarrow f(x, y)=0$, *i.e.*, if its equation remains unaltered when x is replaced by $-x$ and y is replaced by $-y$ throughout. Such a symmetry is often called *rotational symmetry about the origin* because if a curve has this type of symmetry, then the portion of the graph in the third (resp. fourth) quadrant can be obtained by rotating the portion of the graph in the first (resp. second) quadrant through an angle of 180° about the origin.

(e) A curve $f(x, y)=0$ is said to be symmetrical with respect to the line $y=x$ if (y, x) is also a point on the graph whenever (x, y) is a point on the graph, *i.e.*, if $f(y, x)=0 \Leftrightarrow f(x, y)=0$. *If a curve is symmetrical with respect to the line $y=x$, it is enough to sketch the portion of the graph on one side of the line $y=x$ and then complete the graph by taking the reflection of this portion in the line $y=x$.*

Illustrations

1. The curve $y^2-x=0$ is symmetrical with respect to the x-axis.

2. The curve $y-x^2=0$ is symmetrical with respect to the y-axis.

3. The curve $x^2y^2=a^2x^2+b^2y^2$ is symmetri cal with respect to both the axes.

4. The curve $y=x^3$ is symmetrical with respect to the origin.

5. The curve $x^3+y^3=3axy$ is symmetrical with respect to the line $y=x$.

2. Quadrants. Sometimes it is possible to see that graph lies in certain quadrants only, and that no part of the graph lies in certain other quadrants. For example, (i) no portion of the graph of the curve $y^2(4-x)=x^3$ can lie in the second and the third quadrants. (Observe that if $x<0$, then $y^2<0$, which is not possible).

(ii) No portion of the graph of the curve $x^4+y^4=4xy$ can lie in the second and the fourth quadrants (Observe that $x^4+y^4 \geqslant 0$ for all values of x, y so that $4xy$ can never be negative, *i.e.*, x and y cannot be of opposite signs).

3. Range and Domain. The permissible values of x and y can often be seen. For example, (i) Consider the curve $y^2(2a-x)=x^3$, $a>0$. For this curve, $y^2=x^3/(2a-x)$, or $y=\pm x[x/(2a-x)]^{1/2}$. It is obvious that we cannot have $x>2a$ or $x<0$. Therefore the graph must lie in the region bounded by the lines $x=0$ and $x=2a$.

(*ii*) Consider the curve $y(1+x^2)=1$. For this curve, $y=1/(1+x^2)$. It is obvious that, whatever x may be, $y>0$ and $y\leq 1$. Thus the graph must lie in the region bounded by the lines $y=0$ and $y=1$.

4. Points of intersection with the axes of co-ordinates. It is useful to find the points at which the graph of the curve $f(x, y)=0$ intersects the axes of co-ordinates.

To find the points on the x-axis, solve $f(x, y)=0$ and $y=0$ together.

To find the points on the y-axis, solve $f(x, y)=0$ and $x=0$ together.

The points on the graph lying on the axes of co-ordinates act as *gates* through which the graph enters from one quadrant to another. It is useful to find the equations of the tangents to the curve at these points. (Remember that the equation of the tangent(s) to the curve at the origin can be obtained by equating to zero the lowest degree terms in the equation of the curve).

5. Discontinuities. Find the values of x (if any) which give discontinuities. Find the behaviour of the curve near the points of discontinuity.

6. Shape of the curve near the origin. If the curve passes through the origin, the shape of the curve near the origin can be obtained by retaining only the most predominant terms and dropping out the other terms. For example, consider the curve $y=x^2+2x^3-2x^6$. The curve passes through the origin. For small values of x, $x^2+2x^3-3x^6 \approx x^2$, so that the curve approximates to the parabola $y=x^2$ in the neighbourhood of the origin.

7. Behaviour for large x. The approximate shape of the curve for large values of x can be obtained by retaining the terms involving the highest power of x.

8. Minima and Maxima. Examine the curve for critical points (by equating dy/dx to zero). For each critical point find whether the curve has a maximum or a minimum there.

9. Variation of y with respect to x. If it is possible to solve the equation of the curve for y in terms of x, do so, and by examining the sign of dy/dx find in $]-\infty, \infty[$ the intervals in which y is increasing or decreasing. This step is often very important.

10. Intuition and Experience is the most helpful guide for sketching graphs of curves. One may collect all the information detailed out in 1–9 above and yet may not be able to sketch the graph if one does not intelligently apply one's common sense.

On the other hand, all the information asked for in 1–9 above may not actually be necessary in every case. Intuition will often

tell us whether the information gathered upto a certain stage is enough for tracing a curve, or whether some more information is necessary.

The following examples will give us some experience of curve tracing and will enable us to handle the curves given in the problems (or encountered elsewhere) successfully.

Example 31. *Trace the curve* $y=x^3$.

Solution.

1. Since the equation of the curve remains unchanged when x is changed to $-x$ and y is changed to $-y$, therefore the curve is symmetrical about the origin. There are no other symmetries.

2. If $x>0$, then $y>0$, and if $x<0$, then $y<0$. Therefore the graph of the curve lies in the first and the third quadrants only.

3. For each $x \in \mathbf{R}$, y is real. Therefore the domain of the function $y=x^3$ is $\mathbf{R}$.

4. The curve meets the axes of co-ordinates at the origin only.

5. The tangent at the origin is $y=0$.

6. $dy/dx=3x^2$. Since dy/dx vanishes at $x=0$ only, but does not change sign as x passes through 0, therefore y does not possess any extreme values.

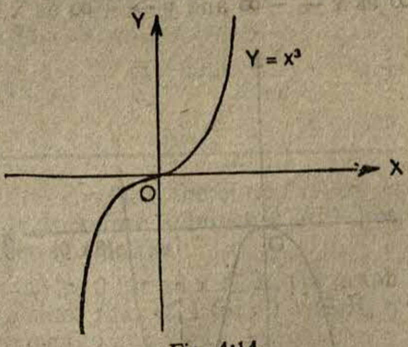


Fig. 4·14.

7. Since $dy/dx>0$ in $]-\infty, 0[$ as also in $]0, \infty[$, therefore y is strictly increasing in $]-\infty, 0]$ as well as in $[0, \infty[$, *i.e.*, in $]-\infty, \infty[$.

8. The tangent at (0, 0) is the line $y=0$.

9. $y \to -\infty$ as $x \to -\infty$, and $y \to +\infty$ as $x \to +\infty$.

The graph of the curve is easily seen to be as shown in Fig. 4.14.

Example 32. *Trace the curve $y=x^2(x-3)$.*

Solution.

1. The curve does not possess any symmetries.

2. If $x<3$, then $y^2<0$. In particular, no portion of the graph lies in the second quadrant.

3. The curve meets the x-axis at points given by $x=0, 3$ *i.e.*, at $(0, 0)$, $(3, 0)$. It meets the y-axis only at the origin.

4. The tangent at the origin is the line $y=0$. Since $y < 0$ for small values of x, therefore it follows that the shape of the graph in the neighbourhood of the origin is as shown in Fig. 4·15.

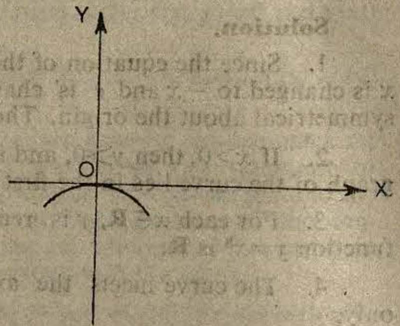


Fig. 4·15.

5. $dy/dx=3x(x-2)$.

Since $dy/dx > 0$ when $x < 0$, therefore, y is increasing in $[-\infty, 0]$.

Since $dy/dx < 0$ in $]0, 2[$, therefore y is decreasing in $[0, 2]$.

Again, since $dy/dx > 0$ if $x > 2$, therefore y is increasing in $]2, \infty[$.

Also, y has a maximum at $(0, 0)$ and a minimum at $(2, -4)$.

6. $y \to -\infty$ as $x \to -\infty$ and $y \to +\infty$ as $x \to +\infty$.

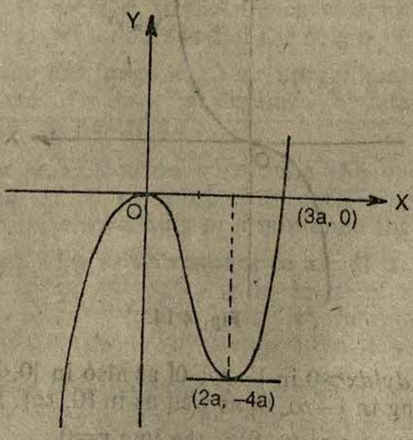


Fig. 4·16.

The graph of the curve can now be seen to be as shown in Fig. 4·16.

Example 33. *Trace the curve* $y=x^3-x^2-2x$.

Solution. 1. The curve does not possess any symmetries.

2. The curve meets the x-axis at the points given by $x^3-x^2-2x=0$, *i.e.*, $x=-1, 0, 2$. Therefore the curve meets the x-axis at the points $(-1, 0)$, $(0, 0)$, $(2, 0)$.

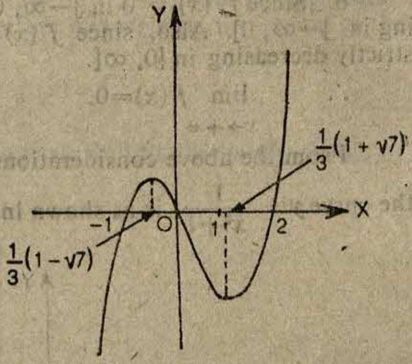


Fig. 4·17.

3. $dy/dx=3x^2-2x-2$,
$=3(x-\alpha)(x-\beta)$,
where α, β are the roots of $3x^2-2x-2=0$, so that $\alpha=\frac{1}{3}(1-\sqrt{7})$, $\beta=\frac{1}{3}(1+\sqrt{7})$, say.

If $x<\alpha$, $dy/dx>0$ so that y in increasing in the interval $]-\infty, \alpha]$.

If $\alpha<x<\beta$, $dy/dx<0$ so that y is decreasing in the interval $[\alpha, \beta]$.

If $x>\beta$, $dy/dx>0$ so that y is increasing in the interval $[\beta, +\infty]$.

Also, y is a maximum at $x=\alpha$, and a minimum at $x=\beta$.

5. $y \to -\infty$ as $x \to -\infty$, and $y \to +\infty$ as $x \to +\infty$.

The graph of the curve is as shown in Fig. 4·17.

Example 34. *Sketch the curves* :

(*a*) $y=\dfrac{1}{x^2+1}$ (*b*) $y=\dfrac{x}{x^2+1}$ (*c*) $y=\dfrac{|x|}{x^2+1}$.

Solution. (*a*) Let $f(x)=\dfrac{1}{x^2+1}$.

1. Since $f(-x)=f(x)$, therefore f is an even function of x. The graph of f is therefore symmetric with respect to the y-axis. There are no other symmetries.

2. Since $f(x)>0$ for all $x \in \mathbf{R}$, the graph lies entirely above the x-axis. Also since $f(x) \leqslant 1$ for all $x \in \mathbf{R}$, therefore no part of the graph lies above the line $y=1$.

3. When $x=0$, $f(x)=1$, therefore $(0, 1)$ lies on the graph. The graph does not meet the axes of co-ordinates at any other point.

4. f is continuous for all $x \in \mathbf{R}$, and therefore there is no break in the graph.

5. $f'(x)=\dfrac{-2x}{1+x^2}$. Since $f'(x)=0$ when $x=0$, therefore the tangent at $(0, 1)$ is parallel to the x-axis. Also $f'(x)$ changes sign

from positive to negative as x passes through 0. Therefore f has a maximum at $x=0$.

6. Since $f'(x) > 0$ in $]-\infty, 0[$, therefore f is strictly increasing in $]-\infty, 0]$. Also, since $f'(x) < 0$ in $]0, \infty[$, therefore f is strictly decreasing in $[0, \infty[$.

$$\therefore \quad \lim_{x \to +\infty} f(x)=0.$$

From the above considerations we find that a rough sketch of the curve $y=\dfrac{1}{x^2+1}$ is as shown in Fig. 4·18 (*a*).

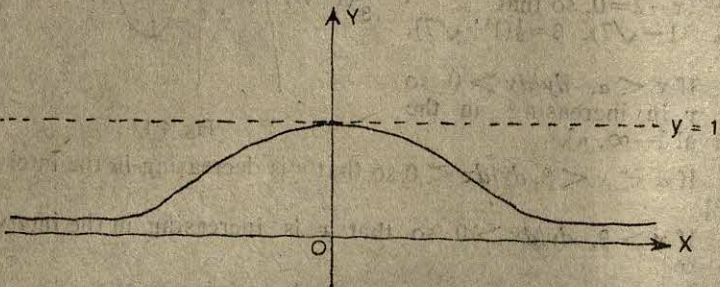


Fig. 4·18 (*a*).

(*b*) Let $g(x)=\dfrac{x}{x^2+1}$.

1. $g(-x)=-g(x)$, therefore g is an odd function of x. The graph of g has rotational symmetry about the origin. There are no other symmetries.

2. Since $g(x)$ has the same sign as x, therefore no part of the graph can lie either in the second quadrant or in the fourth quadrant.

3. Since $g(x)=0$ iff $x=0$, therefore $(0, 0)$ lies on the graph. The graph does not meet the axes at any other point.

4. g is continuous for all $x \in \mathbf{R}$, and therefore there is no break in the graph.

5. $$g'(x)=\frac{1 \cdot (x^2+1)-x \cdot 2x}{(x^2+1)^2}=\frac{1-x^2}{(1+x^2)^2}.$$

$g'(x)=0$ when $x=\pm 1$. Also, as x passes through the value -1, $g'(x)$ changes sign from negative to positive, so that $g(x)$ has a minimum at $x=-1$.

Again, as x passes through the value 1, $g'(x)$ changes sign from positive to negative, so that g has a maximum at $x=1$.

6. The equation of the tangent at the origin is $y=x$. Since $g(x) \leqslant x$ if $x \geqslant 0$ and $g(x) \geqslant x$ if $x < 0$, therefore the graph of g

lies below the line $y=x$ in the first quadrant, and above it in the third quadrant.

7. $g'(x) < 0$ in $]-\infty, -1[$ and $]1, \infty[$. Therefore g is strictly decreasing in each of the intervals $]-\infty, -1]$ and $[1, \infty[$. Also, since $g'(x) > 0$ in $]-1, 1[$, therefore g is strictly increasing in the interval $[-1, 1]$.

8. $\lim\limits_{x \to +\infty} g(x)=0, \quad \lim\limits_{x \to -\infty} g(x)=0.$

From the above considerations we find that a rough sketch of g is as shown in Fig. 4·18 (*b*).

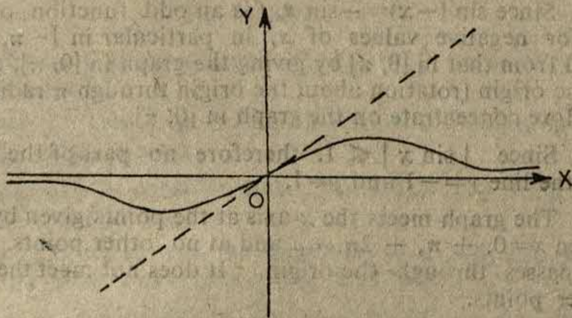


Fig. 4·18 (*b*).

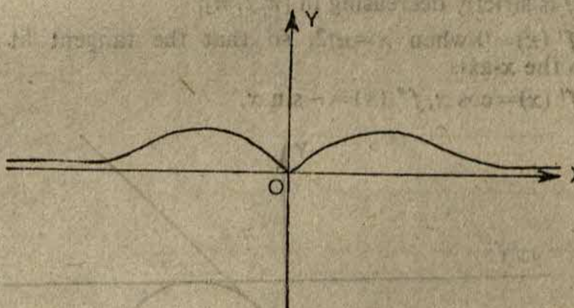


Fig. 4·18 (*c*).

(*c*) Let $h(x)=\dfrac{|x|}{x^2+1}$.

Now $h(x)=\dfrac{x}{x^2+1}=g(x)$ if $x \geqslant 0$.

Also $h(x)=-\dfrac{x}{x^2+1}=-g(x)$ if $x \leqslant 0$.

The graph of h can, therefore, be obtained from that of g by reflecting the portion of the latter lying in the left-hand half of the plane in the x-axis. The graph is, therefore, as shown in Fig. 4·18 (c).

Example 35. *Sketch the graphs of* (a) $y=\sin x$, (b) $y=\cos x$, (c) $y=|\sin x|$, (d) $y=2\sin x$, (e) $y=-2\sin x$.

Solution. (a) Let $f(x)=\sin x$.

1. Since $\sin(x+2\pi)=\sin x$, for all $x \in \mathbf{R}$, therefore f is periodic with period 2π, and consequently it is enough to sketch the graph over one period, say in $[0, 2\pi]$, or in $[-\pi, \pi]$.

2. Since $\sin(-x)=-\sin x$, f is an odd function of x. The graph for negative values of x, in particular in $[-\pi, 0]$, can be obtained from that in $[0, \pi]$ by giving the graph in $[0, \pi]$, a half-turn about the origin (rotation about the origin through π radians). Let us therefore concentrate on the graph in $[0, \pi]$.

3. Since $|\sin x| \leqslant 1$, therefore no part of the graph lies outside the line $y=-1$ and $y=1$.

4. The graph meets the x-axis at the points given by $\sin x=0$, *i.e.*, when $x=0, \pm\pi, \pm 2\pi, \ldots$ and at no other points. In particular it passes through the origin. It does not meet the y-axis at any other points.

5. $f'(x)=\cos x$. Since $f'(x)>0$ in $[0, \pi/2[$ therefore f is strictly increasing in $[0, \pi/2[$. Also, since $f'(x)<0$ in $]\pi/2, \pi[$, therefore f is strictly decreasing in $[\pi/2, \pi]$.

6. $f'(x)=0$ when $x=\pi/2$, so that the tangent at $x=\pi/2$ is parallel to the x-axis

7. $f'(x)=\cos x, f''(x)=-\sin x$.

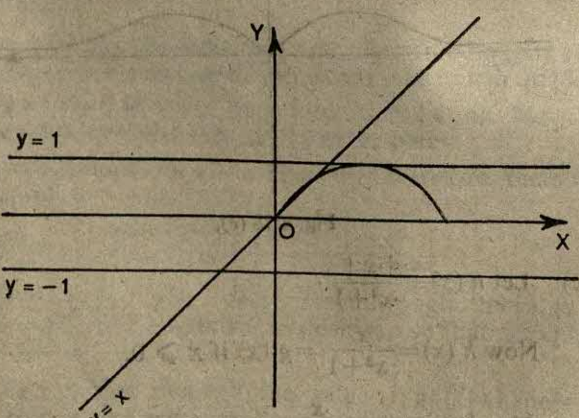


Fig. 4·19 (a).

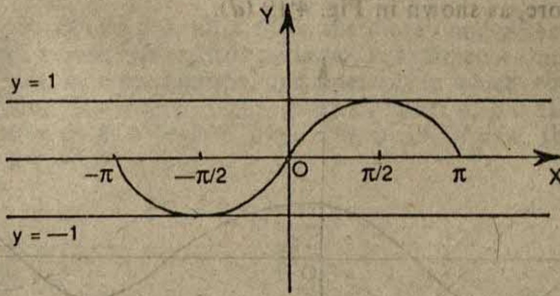


Fig. 4·19 (b).

Since $f'(x)=0$ at $x=\pi/2$, and $f''(\pi/2)<0$, therefore f has a maximum at $x=\pi/2$.

8. Since $f'(x)=1$ at $x=0$, therefore the equation of the tangent at the origin is $y=x$.

From the above considerations, the graph of f in $[0, \pi]$ is as shown in Fig. 4·19 (a).

By using (2), we extend the graph to $[-\pi/2, 0]$ also, and the graph in $[-\pi/2, \pi/2]$ is as shown in Fig. 4·19 (b).

We now use periodicity to get the graph of $y=\sin x$ as shown in Fig. 4·19 (c).

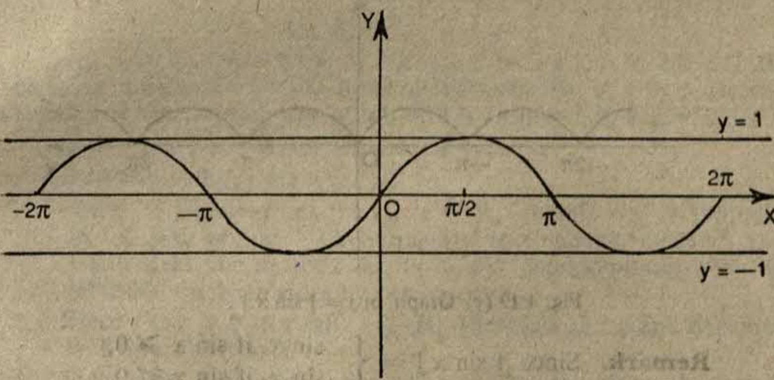


Fig. 4·19 (c).

(b) Since $\cos p=\sin(p+\pi/2)$, therefore the ordinate of $\cos x$ at $x=p$ is the same as that of $\sin x$ at $x=p+\pi/2$ for all $p\in\mathbf{R}$. This means that the graph of $\sin x$ can be obtained from that of $\cos x$ by pushing (translating) the latter forward (parallel to the positive direction of the x-axis) through a distance $\pi/2$. Equivalently, this means that if we push the graph of $y=\sin x$ backwards through a

distance $\pi/2$, we shall get the graph of $y = \cos x$. The graph of $\cos x$ is, therefore, as shown in Fig. 4·19 (*d*).

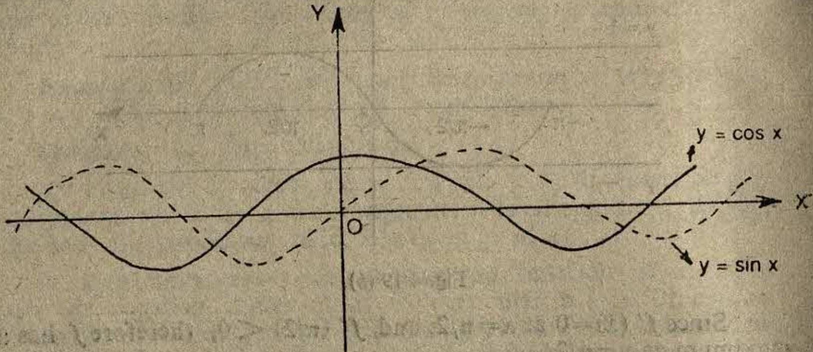


Fig. 4·19 (*d*). Graphs of $y=\cos x$ and $y=\sin x$.

(*c*) Since $|\sin(x+\pi)| = |-\sin x| = |\sin x|$, therefore $|\sin x|$ is periodic and has a period π.

Also, $|\sin x| = \sin x, \ \forall \ x \in [0, \pi]$.

Therefore the curve $y = |\sin x|$ has the same graph as $y=\sin x$ in $[0, \pi]$. Thus to get the graph of $|\sin x|$, we shall draw the graph of $y = \sin x$ in $[0, \pi]$ and repeat it on both sides. The graph is as shown in Fig. 4·19 (*e*).

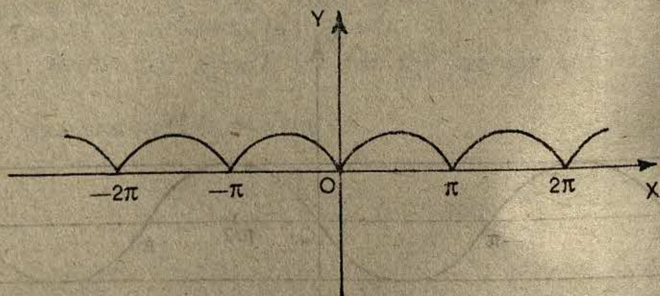


Fig. 4·19 (*e*) Graph of $y = |\sin x|$.

Remark. Since $|\sin x| = \begin{cases} \sin x, \text{ if } \sin x \geqslant 0, \\ -\sin x, \text{ if } \sin x < 0, \end{cases}$

therefore we could have first drawn the graph of $y=\sin x$, and then reflected the portions lying below the x-axis in the x-axis to obtain the graph of $y = |\sin x|$.

(*d*) Let $f(x) = \sin x$, $g(x) = 2 \sin x$.

Since $f(x) = 2g(x)$, $f'(x) = 2g'(x)$,

$f''(x)=2g''(x)$ for all $x \in \mathbf{R}$, therefore f and g behave exactly in the same manner as regards periodicity, symmetries, intersections with the axes, maxima-minima, and intervals in which the functions are increasing/decreasing. Also, every ordinate of g is twice that of the corresponding ordinate for f. The graph of g is, therefore, as shown in Fig. 4·19 (f).

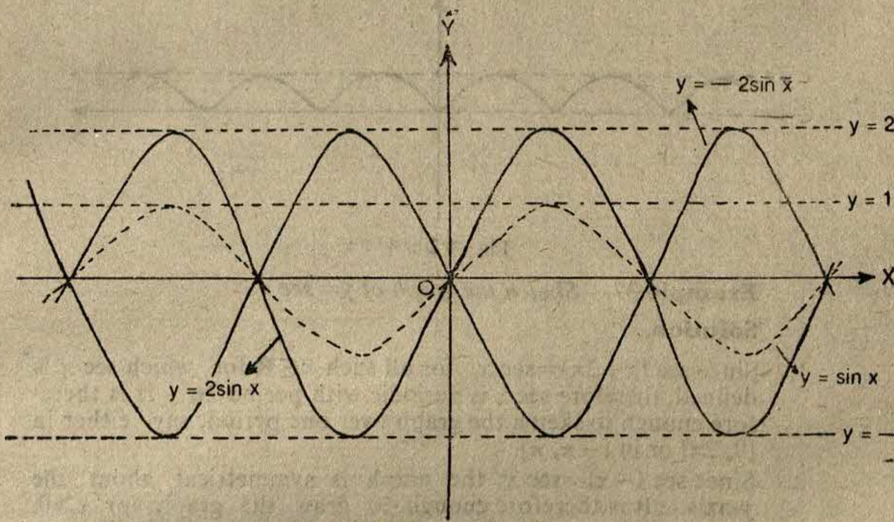


Fig. 4·19 (f).

(e) Let $h(x)=-2 \sin x$. Since $h(x)=-g(x)$ for all $x \in \mathbf{R}$, therefore the graph of h can be obtained from that of g by reflecting the latter in the x-axis. The graph of h is as shown in Fig. 4·19 (f).

Example 36. *Sketch the curve* $y=\sin^2 x$.

Solution. Let $f(x)=\sin^2 x$.

1. Since $f(x+\pi)=f(x)$, for all $x \in \mathbf{R}$, therefore f is periodic and π is a period. Consequently it is enough to sketch the graph over one period, say in $[0, \pi]$. The graph can then be extended on both sides by periodicity.
2. Since $f(x) \geqslant 0$ for all $x \in \mathbf{R}$, therefore the graph does not lie below the x-axis.
3. Since $0 < \sin^2 x \leqslant 1$ for all $x \in \mathbf{R}$, therefore the graph lies entirely in the region bounded by the lines $y=0$ and $y=1$.
4. $f'(x)=2 \sin x \cos x=\sin 2x$, $f''(x)=2 \cos 2x$, $f'(x)=0$ at $x=0, \pi/2, \pi$ (in the interval $[0, \pi]$). Also, $f''(0)>0$, $f''(\pi/2)<0$, $f''(\pi)>0$. Therefore f has a minimum when $x=0$ and π, and a maximum when $x=\pi/2$.
5. Since $f'(x)>0$ in $]0, \pi/2[$, therefore f is strictly increasing in $[0, \pi/2]$. Also since $f'(x)<0$ in $]\pi/2, \pi[$, therefore f is strictly decreasing in $[\pi/2, \pi]$.

6. The tangents at $x=0, \pi/2, \pi$ are $y=0$, $y=1$ and $y=0$ respectively.

From the above considerations we find that the graph of $y=\sin^2 x$ is as shown in Fig. 4·20.

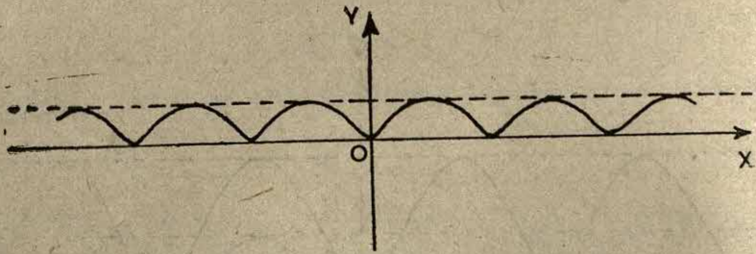


Fig. 4·20.

Example 37. *Sketch the graph of $y=\sec x$.*

Solution.

1. Since $\sec(x+2\pi)=\sec x$, for all such $x\in\mathbf{R}$ for which $\sec x$ is defined, therefore $\sec x$ is periodic with period 2π. It is therefore enough to sketch the graph over one period, say either in $[0, 2\pi]$ or in $[-\pi, \pi]$.
2. Since $\sec(-x)=\sec x$, the graph is symmetrical about the y-axis. It is therefore enough to draw the graph for $x\geqslant 0$. The graph for $x\leqslant 0$ will be obtained by reflecting it in the y-axis. This means that if we first draw the graph in $[0, \pi]$, and reflect it in the y-axis to draw the graph in $[-\pi, 0]$, then we can get the graph in $[-\pi, \pi]$. By using (1) above, we can then draw the graph for other values of x.
3. Since $|\sec x|\geqslant 1$, therefore no part of the graph will lie between the lies $y=-1$ and $y=1$ (*i.e.*, in the region $-1<y<1$).
4. Since $\sec x$ is not defined at $x=\pi/2$, there will be discontinuity in the graph at $x=\pi/2$. Furthermore, $\sec x$ is continuous in $[0, \pi/2[$ and $]\pi/2, \pi]$.
5. $dy/dx=\sec x\tan x>0$ in $]0, \pi/2[$. Therefore $\sec x$ is stricly increasing in $[0, \pi/2[$. Also $y\to+\infty$ as $x\to\pi/2-$ and $y\to-\infty$ as $x\to\pi/2+$.
6. $dy/dx=0$ when $x=0, \pi$ (of course also at $x=-\pi, \pm 2\pi, \ldots\ldots$ also but we are considering only $[0, \pi]$ at the moment). Therefore the tangents at $x=0$ and $x=\pi$ are parallel to the x-axis.
7. $d^2y/dx^2=\sec x\tan^2 x+\sec^3 x=\sec x(\sec^2 x+\tan^2 x)$.

Now $\dfrac{dy}{dx}=0$ and $\dfrac{d^2y}{dx^2}>0$ at $x=0$, therefore the function has a minimum at $x=0$.

Also, since $\frac{dy}{dx}=0$, $\frac{d^2y}{dx^2}<0$ at $x=\pi$, therefore the function has a maximum at $x=\pi$.

From the above observations (1)–(6) we find that the graph in $[0, \pi/2[$ is as shown in Fig. 4·2 (*a*).

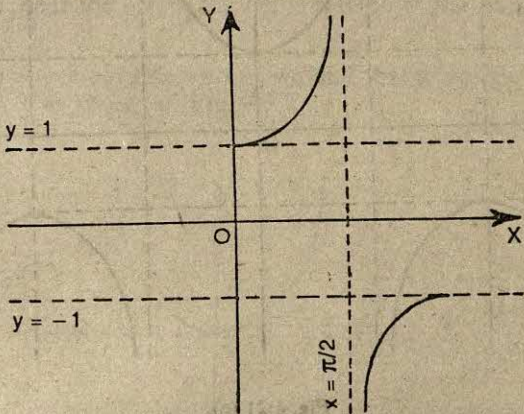


Fig. 4·21 (a).

Using (2), we find that the graph of sec x in $[-\pi, \pi]$ is as shown in Fig. 4·21 (*b*).

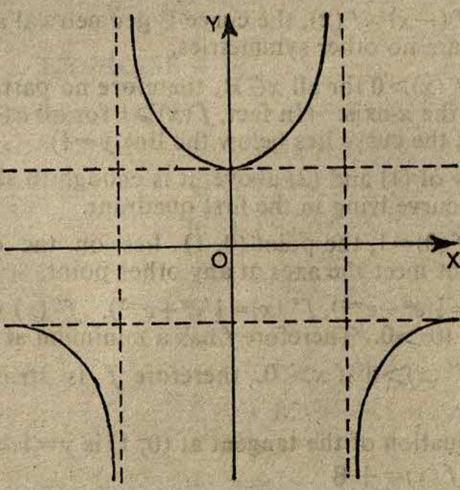


Fig. 4·21 (b).

Using (1), we extend the graph on both sides for $x<-\pi$ and $x>\pi$. The graph is as shown in Fig. 4·21 (c).

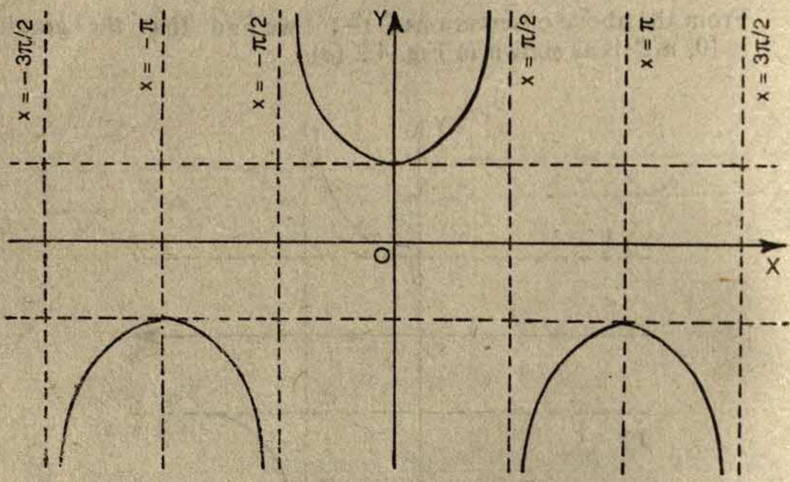


Fig. 4·21 (c).

The graph extends indefinitely on both sides.

Example 38. *Sketch the curve :*

$$y=\tfrac{1}{2}(e^{x}+e^{-x}).$$

Solution. Let $f(x)=\frac{1}{2}(e^{x}+e^{-x})$.

1. Since $f(-x)=f(x)$, the curve is geometrical about the y-axis. There are no other symmetries.
2. Since $f(x)>0$ for all $x\in \mathbf{R}$, therefore no part of the curve lies below the x-axis. (In fact, $f(x)\geqslant 1$ for all $x\in\mathbf{R}$, therefore no part of the curve lies below the line $y=1$).
3. In view of (1) and (2) above, it is enough to sketch the portion of the curve lying in the first quadrant.
4. Since $f(0)=1$, the point (0, 1) lies on the curve. The curve does not meet the axes at any other point.
5. $f'(x)=\frac{1}{2}(e^{x}-e^{-x}), f''(x)=\frac{1}{2}(e^{x}+e^{-x})$. $f'(x)$ vanishes at $x=0$, and $f''(0)>0$. Therefore f has a minimum at $x=0$.
6. Since $f'(x)>0$ if $x>0$, therefore f is strictly increasing in $[0, \infty[$.
7. The equation of the tangent at (0, 1) is $y=1$.
8. $\lim\limits_{x\to+\infty} f(x)=+8$

9. For small $x, f(x)$ is approximately equal to $\frac{1}{2}\left[\left(1+x+\frac{x^2}{2!}+\ldots\ldots\right)+\left(1-x+\frac{x^2}{2!}+\ldots\ldots\right)\right]$ *i.e.*, to $1+\frac{1}{2}x^2$, neglecting higher powers of x.

Therefore near the origin the curve resembles the parabola $y=1+\frac{1}{2}x^2$.

From the above considerations, we find that a rough sketch of the curve is as shown in Fig. 4·22.

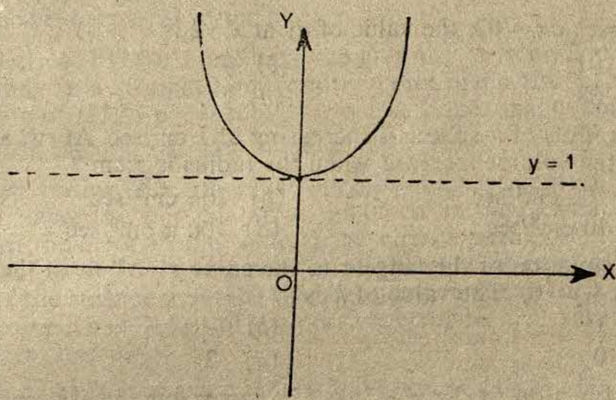


Fig. 4·22.

EXERCISE 4 (*k*)

Sketch the curves :

1. $y=x(y^2-3)$.
2. $y=x(x^2+1)$.
3. $y=(x-2)(x+1)^2$.
4. $y=x^3+3x+1$.
5. $y=x(x-3)^2$.
6. $y(1+x^2)=x^2$.
7. $y(1+x^2)=x^3$.
8. $y=\dfrac{x}{x^2-1}$.
9. (*a*) $y=\tan x$.
   (*b*) $y=|\tan x|$.
10. (*a*) $y=\cot x$.
    (*b*) $y=|\cot x|$.
11. (*a*) $y=\cos^2 x$.
    (*b*) $y=2\cos 2x$.
12. (*a*) $y=\csc x$.
    (*b*) $y=2\csc x$.
13. $y=e^x-1$.
14. $y=1-e^{-x}$.
15. $y=\frac{1}{2}(e^x-e^{-x})$.
16. $y=3e^{2x}+1$.

TEST YOUR UNDERSTANDING IV

For each of the following problems four alternatives are given out of which only one is correct. Put a tick mark (√) against the correct alternative.

1. The distance travelled by a particle in t seconds is given by $s(t)=3t^2-2t+1$. Its velocity at $t=2$ is
   (*a*) 12 m/sec (*b*) 14 m/sec
   (*c*) 10 m/sec (*d*) 9 m/sec.
2. A particle is projected vertically upwards with a velocity 14 m/sec. How high will it go?
   (*a*) 10 m (*b*) 28 m
   (*c*) 100 m (*d*) 56 m.
3. If $y=x^2$, $dx=\cdot 02$, the value of dy at $x=2$ is
   (*a*) ·01 (*b*) 4
   (*c*) ·08 (*d*) ·8.
4. The radius of a sphere is increasing at 1 cm/sec. At what rate will the volume increase when the radius is 5 cm?
   (*a*) 100 cm³/sec (*b*) 10π cm³/sec
   (*c*) 10 cm³/sec (*d*) 100π cm³/sec.
5. The equation of the tangent to the curve $y=x^2-x$ at the origin is $kx+y=0$. The value of k is
   (*a*) 1 (*b*) -1
   (*c*) 0 (*d*) 2.
6. The function $f(x)=x^2-3x+2$ is increasing in $[k, \infty[$. The value of k must be
   (*a*) 0 (*b*) 1
   (*c*) -1 (*d*) 2.
7. The maximum value of $3 \sin (2x+4)$ is
   (*a*) 1 (*b*) 4
   (*c*) 2 (*d*) 3.
8. The function $f(x)=\sin x$ is strictly increasing in
   (*a*) $[0, \pi]$ (*b*) $[-\pi, \pi]$
   (*c*) $[-\pi/2, \pi/2]$ (*d*) $[0, 2\pi]$.
9. The minimum value of $3 \sin x-4 \cos x$ is
   (*a*) 3 (*b*) -4
   (*c*) 5 (*d*) -5.
10. The function $f(x)=x^2$ satisfies all the conditions of Rolle's theorem is
    (*a*) $[2, 4]$ (*b*) $[1, 3]$
    (*c*) $[-2, 1]$ (*d*) $[-2, 2]$.

REVIEW EXERCISE IV

1. The distance $s(t)$ travelled by a particle in t seconds is given by $s(t)=3t^2-2t$. During what time interval is the particle moving in the positive direction ?

2. If $s=1/(1-t)$, $0<t<1$, where the distance s is measured in metres and time is measured in minutes, what is the velocity when $t=\cdot 5$?

3. A ball is dropped from a height of 40 metres. What is its velocity on reaching the ground ?

4. The motion of a particle moving in a straight line is described by $s=2t^3-t$, where the distance s is measured in metres and the time is measured in seconds. What is the acceleration when the velocity is 23 m/sec ?

5. A street light is 6 m above the ground. A man 2 m tall walks away in a straight line from the point under the light at the rate of 1·5 m per second. How fast is the shadow increasing ?

6. Find the approximate value of $(248)^{1/5}$.

7. The radius of a sphere is found by measurement to be 5 cm. If there is a possible error $\pm 0{\cdot}05$ cm in the measurement, find (*a*) the error in the value of the surface area ; (*b*) the error in the value of the volume ; (*c*) the precentage error in the value of the surface area; (*d*) the precentage error in the value of the volume as a result of the possible error in the radius due to measurement.

8. A particle is moving in a straight line such that its distance s metres at any time t seconds is given by $s=\frac{1}{2}t^4-6t^3+15t^2-3$. Obtain its maximum velocity and minimum acceleration. (*D.B.S.S.C.E.*, *1986*)

9. How should a wire 20 cm long be divided into two parts, if one part is to be bent into a circle, the other part is to be bent into a square, and the two plane figures are to have areas the sum of which is maximum ? (*Roorkee Entrance*, *1986*)

10. A firm has a branch store in each of three cities A, B and C. A and B are 320 km apart and C is 200 km from each of them. In order to minimize the time of transportation. it should be located so that the sum of the distances from the godown to each of the cities is a minimum. Where should the godown be built ? (*Roorkee Entrance*, *1981*)

11. Find the maximum and minimum value of the function $40/(3x^4+8x^3-18x^2+60)$. (*Roorkee Entrance*, *1989*)

12. A rectangular box with square base and open top is to be made from 1200 cm^2 of cardboard. Find the maximum possible volume of such a box.

13. Find the dimensions of the right circular cylinder of maximum volume inscribed in a sphere of radius r.

14. Find the maximum value of $a \cos \theta + b \sin \theta$.

15. Find the maximum and minimum values of $a \cos^2 \theta + b \sin^2 \theta$.

16. Find the minimum value of

$$a^2 \sec^2 \theta + b^2 \csc^2 \theta.$$

17. Show that $\dfrac{\ln x}{x}$ is maximum when $x = e$.

18. Find the equations of the normals to the curve $3x^2 - y^2 = 8$ parallel to $x + 3y = 4$. (*D.B.S.S.C.E., 1985*)

19. Tangents are drawn from the origin to the curve $y = \sin x$. Prove that their points of contact lie on $x^2y^2 = x^2 - y^2$. (*Roorkee Entrance, 1986*)

20. In the curve $x^a y^b = k^{a+b}$, prove that the portion of the tangent intercepted between the co-ordinate axes is divided at its point of contact into segments which are in a constant ratio (all segments being positive). (*Roorkee Entrance, 1988*)

SUMMARY

1. If a particle moving in a straight line travels a distance s units in time t units, then its velocity at time t is $\dot{s}$ and acceleration at time t is $\ddot{s}$.
2. Acceleration due to gravity is 9·8 m/sec^2.
3. If a particle is projected with a velocity v in a direction making an angle α with the horizontal, its position (x, y) at time t is given by

$$x = (u \cos \alpha)\, t,\ y = (u \sin \alpha) t - \frac{1}{2} g t^2.$$

4. If $y = f(x)$, then $dy = f'(x)\, dx$.
5. The equation of the tangent at the point (x, y) on the curve $Y = f(x)$ is
$$Y - y = f'(x)\,(X - x),$$
where (X, Y) are the current co-ordinates of any point on the tangent.
6. The equation of the normal at the point (x, y) on the curve $y = f(x)$ is
$$(X - x) + f'(x)\,(Y - y) = 0,$$
where (X, Y) are the current co-ordinates of any point on the normal.
7. **Rolle's Theorem** : Let f be a function defined on $[a, b]$ such that
   (*i*) f is continuous on $[a, b]$,
   (*ii*) derivable in $]a, b[$,
   (*iii*) $f(a) = f(b)$.
   Then there exists a real number c lying between a and b such that $f'(c) = 0$.
8. **Lagrange's mean value theorem** : Let f be a function defined on $[a, b]$ such that
   (*i*) f is continuous on $[a, b]$,
   (*ii*) f is derivable on $]a, b[$.

Then there exists a real number c lying between a and b such that

$$f(b)-f(a)=(b-a)f'(c)$$

9. If f is continuous on $[a, b]$ and $f'(x)>0$, then f is strictly increasing in $[a, b]$.
10. If f is continuous on $[a, b]$ and $f'(x)<0$, then f is strictly decreasing in $[a, b]$.
11. If f is differentiable on $[a, b]$ and has an extremum at $c \in]a, b[$, then $f'(c)=0$.
12. **First derivative test.** Let f be derivable on an open interval I and let $f'(c)=0$ at some point $c \in$ I. If $f'(x)$ changes sign from positive to negative (resp. negative to positive) as x passes through c, then f has a maximum (resp. minimum) at $x=c$.
13. **Second derivative test.** Let f be derivable on an open interval I and let $f'(c)=0$ at some point $c \in$ I. If $f''(c)>0$ (resp. <0), then f has a minimum (resp. maximum) at $x=c$.

HISTORICAL NOTE

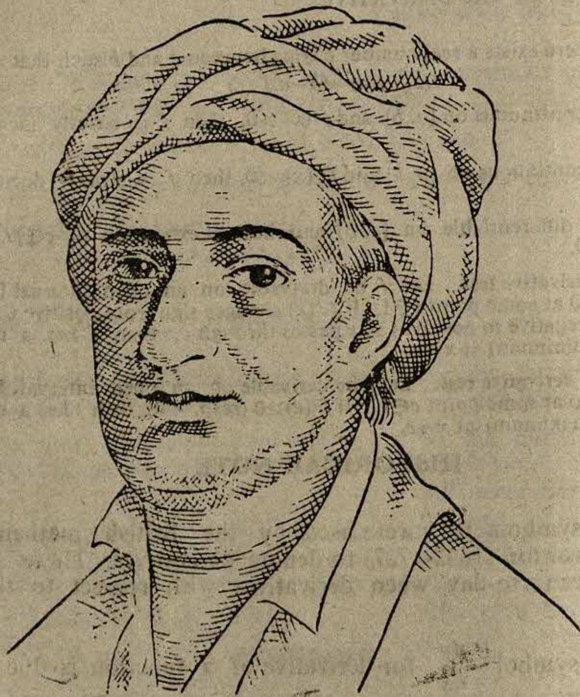
The symbols $\dot{y}$, $\ddot{y}$ were used by the British mathematician Sir Issac Newton (1642-1727) to denote derivatives. These symbols are used even to-day when derivatives with respect to time are required.

The symbol $\dfrac{dy}{dx}$ for derivative of a function is due to the German mathematician Gottfried Wilhelm Leibnitz (1646-1716).

The notation f', f'', f''' for the derivative of a function f is due to the famous French mathematician Joseph Louis Lagrange (1736-1813).

Lagrange's notation is superior to that of Leibnitz in the sense that is emphasizes the important fact that the derivative of a function f is another function f'.

□□



GOTTFRIED WILHELM LEIBNITZ (1646-1716)

Gottfried Wilhelm Leibnitz was born in Leipzig in 1646. When he was only eight years old, he could read Latin and Greek. Mathematics was only one of the many fields in which Leibnitz showed conspicuous genius. Before he was twenty, he had mastered the ordinary textbook knowledge of mathematics, philosophy, theology and law.

Leibnitz is said to have lived not one life but several. As a diplomat, historian, philosopher and mathematician, he did enough in each field to fill one ordinary working life. It was in 1672-73, while he was on a diplomatic mission, first in Paris and then in London, that he exhibited his Calculating Machine to the Royal Society, discovered the fundamental theorem of the Calculus, developed much of his notation in this subject and worked out a number of elementary formulas for differentiation. In 1675, he first used the modern integral sign as a long letter S derived from the first letter of the Latin word summa (sum), to indicate the sum of Cavalieri's indivisibles. His method for finding the *n*th derivative of the product of two functions is still referred to as Leibnitz rule. His notation in Calculus proved to be more convenient than the fluxions of Newton. The theory of determinants is said to have originated with Leibnitz. He also did much in laying the foundation of the theory of envelopes, The last seven years of his life were embittered by the controversy which others had brought upon him and Newton concerning whether he had discovered the Calculus independently of Newton. Infact, each discovered the calculus independently of the other. While Newton's discovery was made first, Leibnitz was the earlier one in publishing results.

CHAPTER 5

Indefinite Integrals

5·1. PRIMITIVES/ANTIDERIVATIVES

In the preceding two chapters we discussed how to differentiate certain functions. This amounted to finding the derived functions of certain given functions. Thus for example, if we are given the function f, where $f(x)=x^2$, then differentiation gives us the derived function f' as $f'(x)=2x$. Thus differentiating f, we get f'. A natural question now arises : can we invert this process ? In other words, given a function f, can we find another function F such that $F'(x)=f(x)$, for all x in the domain of F, so that the given function f now becomes the derived function of F. For example, if f, where $f(x)=2x$, is the given function, and if we take $F(x)=x^2$, then we know that $F'(x)=f(x)$. Given f, the process of finding F (whenever we can) such that $F'(x)=f(x)$ is called *integration*. It must have been clear to you by now that *differentiation* and *integration* are *inverse* processes in the sense that if by differentiating a function f we get the function g, then by integrating g we can recover f. It is our object to describe the process of integration in this chapter. We begin by introducing a term that plays the same role in integration as the term *derived function* plays in differentiation. However, before doing that, let us remark that yielding to the prevalent mathematical slang, we shall use the symbol $f(x)$ for f, and similarly $f'(x)$ for f' etc. However, the context will make it clear whether we have the function f in mind or its image $f(x)$ at x.

Definition 5·1. *Given a function $f(x)$, a function $F(x)$ such that*

$$F'(x)=f(x)$$

(in case its exists), is called a **primitive (or antiderivative)** of $f(x)$ and is written as

$$F(x)=\int f(x)\, dx.$$

For example, both $\sin x$ and $\sin x+1$ are primitives of $\cos x$. A function need not have a primitive, but if it has, then it has a whole lot of them as shown below.

Theorem 5·1. *If $F(x)$ is a primitive of a function $f(x)$, then $F(x)+C$ is also a primitive of $f(x)$, where C is any constant.*

Proof. Since $F(x)$ is a primitive of $f(x)$, we have,

$$F'(x)=f(x).$$

$$\text{Now, } \frac{d}{dx}[F(x)+C]=F'(x)+0$$
$$=F'(x)=f(x).$$

This shows that $F(x)+C$ is also a primitive of $f(x)$ whatever be the value of the constant C.

Theorem 5·2. *Any two primitives of a function differ by a constant.*

Proof. Suppose $f_1(x)$ ane $f_2(x)$ are two primitives of a function $f(x)$. Then

$$f'_1(x)=f(x),$$

and
$$f'_2(x)=f(x).$$

$$\text{Now, } \frac{d}{dx}[f_1(x)-f_2(x)]=f'_1(x)-f'_2(x),$$
$$=f(x)-f(x)=0.$$

$$\therefore \quad f_1(x)-f_2(x)=C,$$

where C is a constant.

Remarks 1. The above theorem says in effect that if $f_1(x)$ and $f_2(x)$ are two primitives of $f(x)$, then $f_1(x)=f_2(x)+C$ for $C \in \mathbf{R}$.

2. It follows from the above theorem that primitive of a function, if it exists, is not unique. $\int f(x)\,dx$ thus denotes a whole class of functions, that is, the class of *all* primitives of $f(x)$, called the *general primitive* of $f(x)$. The adjective *general* will normally be dropped.

Thus
$$\int f(x)\,dx=F(x)+C,$$

where $F(x)$ is a primitive of $f(x)$ and C is any constant, known as the *constant of integration*. In practice, however, the constant of integration is usually omitted and $\int f(x)\,dx$ stands for any one primitive of $f(x)$. Because of the indefiniteness of the constant C, the general primitive of $f(x)$ is referred to as the *indefinite integral of* $f(x)$ or simply the *integral of* $f(x)$. The process of determining the primitives of a function is called *integration* and the function to be integrated is called the *integrand*.

3. The notation $\int f(x)\,dx$ is due to Leibnitz. 'dx' indicates that x is the independent variable and that the integration is to be performed with respect to x. When the independent variable is different from x, we write that variable in place of x. The symbol '$\int$' resembling an elongated letter 'S' is the first letter of the Greek word 'summa'. In fact, the subject of integration first arose in connection

with finding the areas of some plane regions where one was confronted with the limit of a certain sum when the number of terms in the sum tended to infinity and each term tended to zero. The literal meaning of the phrase 'to integrate' is 'to find the sum of'. It was observed much later that integration and differentiation are inverse operations.

In this chapter, we shall use our knowledge of differentiation to determine integrals of some standard functions. Some standard methods for finding integrals will also be discussed. The summation aspect of integration and its applications to finding areas of plane regions will be considered in the next chapter.

Theorem 5·3. *If a function $f(x)$ has a primitive, then*

$$\frac{d}{dx}\left[\int f(x)\,dx\right]=f(x).$$

Proof. Let $F(x)$ be a primitive of $f(x)$, so that

$$F'(x)=f(x),$$

and

$$\int f(x)\,dx=F(x)+C.$$

Then
$$\frac{d}{dx}\left[\int f(x)\,dx\right]=\frac{d}{dx}\left[F(x)+C\right]$$
$$=F'(x)+0,$$
$$=F'(x)=f(x).$$

Theorem 5·4. *If the function $f(x)$ has a primitive, then for all $k \in \mathbf{R}$,*

$$\int (kf)(x)\,dx=k\int f(x)\,dx.$$

Proof. Suppose $f(x)$ is a function which has a primitive $F(x)$ and that $k \in \mathbf{R}$.

Thus (omitting the constant of integration)

$$\int f(x)\,dx=F(x),$$

where

$$F'(x)=f(x).$$

Now,
$$\frac{d}{dx}(kF(x))=k\frac{d}{dx}\left[F(x)\right]$$
$$=k.F'(x)=k.f(x)$$

$\therefore$ by definition,

$$\int k\,f(x)\,dx=k.\,F(x),$$
$$=k\int f(x)\,dx.$$

This proves the theorem.

Remark. Observe that both $\int (k f)(x)\, dx$ and $k\int f(x)\, dx$ represent families of functions. The equality between them is to be understood as an equality between two sets; do not confuse the equality here to be an equality between two functions. A similar remark applies to the next theorem also.

Theorem 5·5. *If both $f_1(x)$ and $f_2(x)$ have primitives, then*

$$\int (f_1(x) \pm f_2(x))\, dx = \int f_1(x)\, dx \pm \int f_2(x)\, dx.$$

Proof. Let $f_1(x)$ and $f_2(x)$ have primitives $F_1(x)$ and $F_2(x)$, respectively. Then $F'_1(x) = f_1(x)$ and $F'_2(x) = f_2(x)$.

$$\text{Now, } \frac{d}{dx}\Big[\, F_1(x) \pm F_2(x) \Big] = F'_1(x) \pm F'_1(x),$$

$$= f_1(x) \pm f_2(x)$$

$\therefore$ by definition,

$$\int \Big[f_1(x) \pm f_2(x) \Big]\, dx = F_1(x) \pm F_2(x),$$

$$= \int f_1(x)\, dx \pm \int f_2(x)\, dx,$$

which proves the result.

Corollary. If $f_1(x), \ldots\ldots, f_n(x)$ have primitives, then

$$\int \Big[f_1(x) \pm \cdots\cdots \pm f_n(x) \Big]\, dx = \int f_1(x)\, dx \pm \ldots \pm \int f_n(x)\, dx.$$

Example 1. *Evaluate :*

(*i*) $\int x\, dx$ (*ii*) $\int e^x\, dx$

(*iii*) $\int \frac{1}{x^2}\, dx,\ x \neq 0$ (*iv*) $\int 7\, x^6\, dx$

(*v*) $\int (sec^2\, x + e^x)\, dx.$

Solution. (*i*) Since $\frac{d}{dx}\left(\frac{x^2}{2}\right) = \frac{1}{2}, 2x = x,$

$\therefore \int x\, dx = \frac{x^2}{2} + C,$

where C is constant.

(*ii*) Since $\frac{d}{dx}(e^x) = e^x,$

therefore $$\int e^x\, dx = e^x + C,$$

where C a constant,

(*iii*) Since $\frac{d}{dx}\left(-\frac{1}{x}\right) = \frac{1}{x^2}$,

therefore $$\int \frac{1}{x^2}\, dx = -\frac{1}{x} + C,$$

where C is a constant.

(*iv*) Now, $\int 7\, x^6\, dx = 7\int x^6\, dx.$

Also, since $\frac{d}{dx}\left(\frac{x^7}{7}\right) = \frac{1}{7}.\ 7.\ x^{7-1} = x^6$,

$$\therefore \qquad \int x^6\, dx = \frac{x^7}{7} + C.$$

Hence $7\int x^6\, dx = 7\left(\frac{x^7}{7} + C\right) = x^7 + C'.$

where C is a constant.

(*v*) Now, $\int (\sec^2 x + e^x)\, dx = \int \sec^2 x\, dx + \int e^x\, dx,$

Also, $$\frac{d}{dx}(\tan x) = \sec^2 x,$$

$$\therefore \qquad \int \sec^2 x\, dx = \tan x + C_1,$$

where C_1 is a constant,

and $$\int e^x\, dx = e^x + C_2$$

where C_2 is a constant.

$$\therefore \qquad \int \sec^2 x\, dx + \int e^x\, dx = \tan x + e^x + C$$

where C is a constant.

Differentiating $\tan x + e^x + C$, we get $\sec^2 x + e^x$, which is the integrand. Hence our solution is correct. (You should always check your answers in this way.)

EXERCISE 5 (*a*)

Evaluate each of the following :

1. $\int x^4\, dx$ 2. $\int \sqrt{x}\, dx$

3. $\int \cos x\, dx$ 4. $\int \sec x \tan x\, dx.$

5. $\int 3x^4\, dx$ 6. $\int \left(x+\frac{1}{x} \right) dx$

7. $\int (x+1)^2\, dx$ 8. $\int (e^x+\sin x)\, dx$

9. $\int \left(\frac{2}{x}+\sqrt{x}+5\sec^2 x \right) dx$ 10. $\int (x^5-4x^4+5)\, dx$

11. $\int \frac{4x^2+3x+2}{x^3}\, dx.$

5·2. INTEGRATION OF FUNCTIONS IN STANDARD FORM

In this section, we shall obtain formulae for determining primitives of certain functions in the standard form.

5·2·1. Integration of the Power Function

Theorem 5·6. *If $f(x)=x^n$ where $n\neq-1$, then*

$$\int x^n\, dx=\frac{x^{n+1}}{n+1}+C, \textit{ C being a constant.}$$

Proof. Since $\frac{d}{dx}\left(\frac{x^{n+1}}{n+1}\right)=\frac{(n+1)}{n+1}, x^{n+1-1}=x^n,$

therefore $$\int x^n\, dx=\frac{x^{n+1}}{n+1}+C,$$

where C is a constant.

Thus we have the following :

Rule. To obtain the primitive of x^n, $n\neq-1$, increase the index of x by unity and divide by the new increased index.

Example 2. *Evaluate* :

(*i*) $\int x^9\, dx$ (*ii*) $\int x^{2/3}\, dx$ (*iii*) $\int x^{-3/4}\, dx.$

Solution. (*i*) $$\int x^9\, dx=\frac{x^{9+1}}{9+1}+C,$$
$$=\frac{x^{10}}{10}+C.$$

(*ii*) $$\int x^{2/3}\, dx=\frac{x^{(2/3)+1}}{\frac{2}{3}+1}+C=\tfrac{3}{5}x^{5/3}+C.$$

(*iii*) $$\int x^{-3/4}\, dx=\frac{x^{(-3/4+1)}}{-\frac{3}{4}+1}+C=4x^{1/4}+C$$

5·2·2. Integration of the power function with exponent -1.

Consider the function $f(x)=\frac{1}{x+a}$, $x\neq-a$

Since $\frac{d}{dx}[ln(|x+a|)]=\frac{1}{x+a}$,

$$\therefore \quad \int \frac{1}{x+a}\,dx=ln(|x+a|)+C.$$

Corollary. By putting $a=0$ in the above result, we have

$$\int \frac{1}{x}\,dx=ln|x|+C \;(x\neq 0).$$

Example 3. *Evaluate* $\int \frac{1}{x+7}\,dx$.

Solution. Since $\frac{d}{dx}[ln(|x+7|)]=\frac{1}{x+7}$

$$\therefore \quad \int \frac{1}{x+7}\,dx=ln(|x+7|)+C.$$

5·2·3. Integrals Involving Trigonometric Functions

(*i*) $\int \sin x\,dx=-\cos x+C$

$\because \quad \frac{d}{dx}(-\cos x)=-(-\sin x)=\sin x.$

(*ii*) $\int \cos x\,dx=\sin x+C$

$\because \quad \frac{d}{dx}(\sin x)=\cos x.$

(*iii*) $\int \sec^2 x\,dx=\tan x+C,$

$\because \quad \frac{d}{dx}(\tan x)=\sec^2 x.$

(*iv*) $\int \csc^2 dx=-\cot x+C,$

$\because \quad \frac{d}{dx}(-\cot x)=-(-\csc^2 x)=\csc^2 x.$

(*v*) $\int \sec x \tan x\,dx=\sec x+C,$

$\because \quad \frac{d}{dx}(\sec x)=\sec x \tan x.$

(*vi*) $\int \csc x \cot x\,dx=-\csc x+C,$

$\because \quad \frac{d}{dx}(\csc x)=-\cot x \csc x.$

Example 4. *Evaluate* $\int \frac{\cos x}{\sin^2 x}\, dx.$

Solution. $\int \frac{\cos x}{\sin^2 x}\, dx = \int \frac{\cos x}{\sin x} \cdot \frac{1}{\sin x}\, dx,$

$$= \int \cot x \csc x\, dx = -\csc x + C.$$

Example 5. *Evaluate* $\int \frac{1-\cos^2 x}{1+\cos x}\, dx$

Solution. $\int \frac{1-\cos^2 x}{1+\cos x}\, dx = \int \frac{(1-\cos x)(1+\cos x)}{1+\cos x}\, dx,$

$$= \int (1-\cos x)\, dx = x - \sin x + C.$$

5·2·4. Integration of the Exponential Function

We have already seen that $\int e^x\, dx = e^x + C.$

Now, consider the function $f(x) = a^x,\ a > 0.$

Since $\frac{d}{dx}\left(\frac{a^x}{\ln a}\right) = \frac{a^x \cdot \ln a}{\ln a} = a^x,$

$\therefore \quad \int a^x\, dx = \frac{1}{\ln a} \cdot a^x + C.$

Example 6. *Evaluate* $\int 10^x\, dx.$

Solution. Since $\frac{d}{dx}\left(\frac{1}{\ln 10} \cdot 10^x\right) = \frac{10^x \cdot \ln 10}{\ln 10} = 10^x,$

$\because \quad \int 10^x\, dx = \frac{10^x}{\ln 10} + C.$

5·2·5. The Inverse Trigonometric Functions as Primitives

(*i*) $\displaystyle\int \frac{\mathbf{dx}}{\sqrt{(1-\mathbf{x}^2)}} = \mathbf{\sin^{-1} x + C}$ **or** $\mathbf{-\cos^{-1} x + C},$

$\because \quad \frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{(1-x^2)}} = \frac{d}{dx}(-\cos^{-1} x).$

Remark. Each of the answers $\sin^{-1} x + C$ and $-\cos^{-1} x + C$ is as good as the other because each can be reduced to the other by making an adjustment of $\pm\pi/2$ in the value of C.

(*ii*) $\displaystyle\int \frac{\mathbf{dx}}{1+\mathbf{x}^2} = \mathbf{\tan^{-1} x + C}$ **or** $\mathbf{-\cot^{-1} x + C},$

$\because \quad \frac{d}{dx}(\tan^{-1} x) = \frac{1}{1+x^2} = \frac{d}{dx}(-\cot^{-1} x).$

(*iii*) $\int \frac{dx}{|x| \sqrt{(x^2-1)}} = \sec^{-1} x + C$ or $-\csc^{-1} x + C$,

$\because \quad \frac{d}{dx}(\sec^{-1} x) = \frac{1}{|x| \sqrt{(x^2-1)}} = \frac{d}{dx}(-\csc^{-1} x)$.

Example 7. *Evaluate* $\int \left(\frac{5}{1+x^2} - \frac{7}{\sqrt{(1-x^2)}} \right) dx$.

Solution. $\int \left(\frac{5}{1+x^2} - \frac{7}{\sqrt{(1-x^2)}} \right) dx$

$$= \int \frac{5}{1+x^2}\, dx - \int \frac{7}{\sqrt{(1-x^2)}}\, dx,$$

$$= 5 \int \frac{1}{1+x^2}\, dx - 7 \int \frac{dx}{\sqrt{(1-x^2)}},$$

$$= 5 \tan^{-1} x - 7 \sin^{-1} x + C.$$

EXERCISE 5 (*b*)

Evaluate each of the following :

1. $\int x^{11}\, dx$.
2. $\int x^{-3/7}\, dx$.
3. $\int \frac{1}{\sqrt{x}}\, dx$.
4. $\int \frac{1}{x^5}\, dx$.
5. $\int (x^4 - 2x + 5)\, dx$.
6. $\int \left(x^3 - \frac{1}{x^3} \right) dx$.
7. $\int (2x+3)^2\, dx$.
8. $\int (x+1)(x+2)\, dx$.
9. $\int (2x-3)(5x+4)\, dx$.
10. $\int \frac{x^5 + \sqrt{x} + 2}{x^{\frac{1}{3}}}\, dx$.
11. $\int \frac{1}{x-5}\, dx$.
12. $\int \left(x^3 + \sqrt{x} + \frac{1}{x} \right) dx$.
13. $\int \frac{x+1}{x^2+1}\, dx$.
14. $\int \left(\sec^2 x + \frac{1}{x+3} \right) dx$.
15. $\int \left(\frac{1}{\sqrt{x}} - \frac{5}{x+9} + \frac{1}{x-11} \right) dx$.
16. $\int (3 \sin x + 8 \cos x)\, dx$.
17. $\int (11 \cos x + 5 \sec^2 x - 4)\, dx$.
18. $\int \left(15 \csc x \cot x + \frac{1}{7} \sec^2 x + 9 \sin x \right) dx$.

19. $\int \frac{5}{\cos^2 x}\, dx.$

20. $\int \frac{4}{\sin^2 x}\, dx.$

21. $\int \tan^2 x\, dx.$

22. $\int \cot^2 x\, dx.$

23. $\int \frac{\cos^2 x}{1-\sin x}\, dx.$

24. $\int \frac{1-\cos x}{\sin^2 x}\, dx.$

25. $\int \left(\frac{\cos^2 x}{1+\sin x}\right) dx.$

26. $\int 5^x\, dx.$

27. $\int (e^x+2^x)\, dx.$

28. $\int (1+2e^x)\, dx.$

29. $\int e^{x+7}\, dx.$

30. $\int (3^x+7\sec^2 x+5e^x)\, dx.$

31. $\int \frac{11}{\sqrt{(1-x^2)}}\, dx.$

32. $\int \left(x^5-\frac{5}{|x|\sqrt{(x^2-1)}}\right) dx.$

33. $\int \left(\frac{5}{x^4}+6\sec^2 x+\frac{1}{|x|\sqrt{(x^2-1)}}\right) dx.$

34. $\int \frac{x^2+7}{x^2+1}\, dx.$

35. $\int \frac{x^4+x^2+4}{x^2+1}\, dx.$

5·3. INTEGRATION BY SUBSTITUTION

The method of integration by substitution consists in changing the independent variable of the integrand to another variable by making a suitable substitution so that the transformed function is easily integrable.

Theorem 5·7. *Let $f(x)$ be a function which has a primitive. If $x=\phi(t)$ is a derivable function of t, then*

$$\int f(x)\, dx=\int f(\phi(t)\phi'(t))\, dt.$$

Proof. Suppose $F(x)$ is a primitive of $f(x)$. Then $F'(x)=f(x)$.

Now $\frac{d}{dt}[F(x)]=\frac{d}{dx}[F(x)]\cdot\frac{dx}{dt}$,

$$=F'(x)\cdot\frac{dx}{dt},$$

$$=f(x)\cdot\frac{dx}{dt},$$

$$=f(\phi(t))\cdot\phi'(t).$$

$$\therefore \qquad F(x)=\int f(\phi(t))\phi'(t)\, dt.$$

Thus $\int f(x)\,dx = F(x) = \int f(\phi(t))\phi'(t)\,dt.$

This proves the theorem.

Corollary 1. $\int [f(x)]^n f'(x)\,dx = \frac{1}{n+1}[f(x)]^{n+1} + C.$

Proof. Let $f(x) = t$.

Then $\frac{dt}{dx} = f'(x).$

Now, $\int [f(x)]^n f'(x)\,dx = \int t^n \,.\, dt,$

$$= \frac{t^{n+1}}{n+1} + C,$$

$$= \frac{[f(x)]^{n+1}}{n+1} + C.$$

Corollary 2. $\int \frac{f'(x)}{f(x)}\,dx = ln\,(\,|f(x)|\,) + C,$

Proof. Let $f(x) = t$.

Then $\frac{dt}{dx} = f'(x).$

Now, $\int \frac{f'(x)}{f(x)}\,dx = \int \frac{dt}{t} = ln\,|\,t\,| + C,$

$$= ln\,(\,|f(x)|\,) + C,$$

Corollary 3. If $F(t)$ is a primitive of $f(t)$, then

$$\int f(ax+b)\,dx = \frac{1}{a}F(ax+b) + C.$$

Proof. Suppose $ax+b = t$.

Then $a\,dx = dt.$

$$\int f(ax+b)\,dx = \frac{1}{a}\int f(t)\,dt,$$

$$= \frac{1}{a}F(t) + C,$$

$$= \frac{1}{a}F(ax+b) + C.$$

For suitable choices of the function f, we have the following special cases :

(i) $\int (ax+b)^n\,dx = \frac{1}{a} \cdot \frac{(ax+b)^{n+1}}{n+1} + C.$

(ii) $\int \frac{dx}{ax+b} = \frac{1}{a} \ln |ax+b| + C.$

(iii) $\int e^{ax+b}\, dx = \frac{1}{a} \cdot e^{ax+b} + C.$

(iv) $\int a^{px+q}\, dx = \frac{1}{p} \cdot \frac{a^{px+q}}{\ln a} + C,\ a>0.$

(v) $\int \sin (ax+b)\, dx = -\frac{1}{a} \cos (ax+b) + C.$

(vi) $\int \cos (ax+b)\, dx = \frac{1}{a} \sin (ax+b) + C.$

(vii) $\int \sec^2 (ax+b)\, dx = \frac{1}{a} \tan (ax+b) + C.$

(viii) $\int \csc^2 (ax+b)\, dx = -\frac{1}{a} \cot (ax+b) + C.$

(ix) $\int \sec (ax+b) \tan (ax+b)\, dx = \frac{1}{a} \sec (ax+b) + C.$

(x) $\int \csc (ax+b) \cot (ax+b)\, dx = -\frac{1}{a} \csc (ax+b) + C.$

Remark. In the beginning, however, it will be instructive to evaluate integrals of the above type by actually making the substitution (and not by using the formulae) as illustrated in examples that follow :

Example 8. *Evaluate* $\int (x^2+3) \cdot 2x\, dx.$

Solution. Let $x^2+3=t.$

Then $2x\, dx = dt.$

$\therefore \quad \int (x^2+3)\, 2x\, dx = \int t\, dt = \frac{t^2}{2} + C = \frac{1}{2} (x^2+3)^2 + C$

Example 9. *Evaluate* $\int \cos (2x+3)\, dx.$

Solution. Let $2x+3=t$, Then $2\, dx = dt$

$$\therefore \quad \int \cos (2x+3)\, dx = \frac{1}{2} \int \cos t\, dt = \frac{1}{2} \sin t + C,$$

$$= \frac{1}{2} (\sin (2x+3)) + C.$$

Example 10. *Evaluate* $\int \frac{4x}{2x^2+7}\, dx.$

Solution. Let $2x^2+7=t$. Then $4x\,dx=dt$.

$$\therefore \quad \int \frac{4x\,dx}{2x^2+7}=\int \frac{dt}{t}=ln\mid t\mid+C,$$

$$=ln\,(2x^2+7)+C,$$

since $\mid 2x^2+7\mid=2x^2+7$ for all $x\in\mathbb{R}$.

Example 11. *Evaluate* $\int sin^2 x\, cos\, x\, dx$.

Solution. Let $\sin x=t$. Then $\cos x\,dx=dt$.

$$\therefore \quad \int \sin^2 x \cos x\; dx=\int t^2\; dt=\frac{1}{3}\sin^3 x+C.$$

Example 12. *Evaluate* $\int (4x+5)^6\, dx$.

Solution. Let $4x+5=t$. Then $4\,dx=dt$.

$$\therefore \quad \int (4x+5)^6\,dx=\int \frac{t^6}{4}\;dt=\frac{1}{4}\int t^6\,dt,$$

$$=\frac{1}{4}\cdot\frac{t^{6+1}}{6+1}+C$$

$$=\frac{1}{28}\cdot t^7+C=\frac{1}{28}\,(4x+5)^7+C.$$

Example 13. *Evaluate* $\int csc^2\,(5x+4)\,dx$.

Solution. Let $5x+4=t$. Then $5\,dx=dt$.

$$\therefore \quad \int \csc^2\,(5x+4)\,dx=\frac{1}{5}\int \csc^2\,t\;dt=-\frac{1}{5}\cot t+C,$$

$$=-\frac{1}{5}\cot\,(5x+4)+C.$$

Example 14. *Evaluate* $\int \frac{4x}{5x+3}\,dx$.

Solution. Let $5x+3=t$. Then $5\,dx=dt$.

$$\therefore \quad \int \frac{4x\,dx}{5x+3}=\int \frac{4\,(t-3)/5}{t}\,\frac{dt}{5}$$

$$=\frac{4}{25}\int \frac{t-3}{t}\;dt,$$

$$=\frac{4}{25}\int\left(1-\frac{3}{t}\right)dt,$$

$$=\frac{4}{25}\left[\int dt-3\int\frac{dt}{t}\right].$$

$$=\frac{4}{25}\left[(5x+3)-3\ ln\ (\ |\ 5x+3\ |\ )\right]+C.$$

Example 15. *Evaluate* $\int \frac{1}{1-\cos x}\,dx.$

Solution. We have, $\int \frac{1}{1-\cos x}\,dx=\frac{1}{2}\int \frac{1}{\sin^2(x/2)}\,dx,$

$$=\frac{1}{2}\int \csc^2\left(\frac{x}{2}\right)dx.$$

Let $\frac{x}{2}=t$. Then $\frac{1}{2}\,dx=dt$.

$$\therefore\quad \frac{1}{2}\int \csc^2\left(\frac{x}{2}\right)dx=\int \csc^2 t\,dt,$$

$$=-\cot t+C,$$

$$=-\cot\left(\frac{x}{2}\right)+C.$$

5·3·1. Integration of Some More Standard Trigonometric Functions

(*i*) $\int \mathbf{tan\ x\ dx=ln\ |\ sec\ x\ |+C.}$

Let $\cos x=t.$

Then $-\sin x\,dx=dt.$

$$\therefore \int \tan x\,dx=\int \frac{\sin x\,dx}{\cos x}=-\int \frac{dt}{t}=-ln\ |\ t\ |+C,$$

$$=ln\ \frac{1}{|\ t\ |}+C,$$

$$=ln\ |\sec x\ |+C.$$

(*ii*) $\int \mathbf{cot\ x\ dx=ln\ |\ sin\ x\ |+C.}$

Let $\sin x=t,$

Then $\cos x\,dx=dt,$

$$\therefore \int \cot x\,dx=\int \frac{\cos x\,dx}{\sin x}=\int \frac{dt}{t}=ln\ |\ t\ |+C,$$

$$=ln\ |\sin x\ |+C.$$

(*iii*) $\int \mathbf{csc\ x\ dx=ln\ |\ tan\ x/2\ |+C.}$

Now $\int \csc x\,dx=\int \frac{1}{\sin x}\,dx,$

$$=\int \frac{1}{2 \sin (x/2) \cos (x/2)}\, dx,$$

$$=\frac{1}{2}\int \frac{\sec^2 (x/2)}{\tan (x/2)}\, dx$$

Let $\tan (x/2)=t$. Then $\frac{1}{2} \sec^2 (x/2)\, dx=dt$.

$$\therefore \quad \int \csc x\, dx=\int \frac{dt}{t}=ln \mid t \mid +C,$$

$$=ln \mid \tan (x/2) \mid +C.$$

(*iv*) $\int$ **sec x dx = ln | sec x + tan x | + C,**

Now $\int \sec x\, dx=\int \frac{\sec x (\sec x+\tan x)}{\sec x+\tan x}\, dx$

Let $\qquad \sec x+\tan x=t.$

Then $\qquad (\sec x \tan x+\sec^2 x)\, dx=dt.$

$$\therefore \quad \int \sec x\, dx=\int \frac{dt}{t}=ln \mid t \mid +C,$$

$$=ln \mid \sec x+\tan x \mid +C.$$

Example 16. *Evaluate* $\int \csc (3x+4)\, dx$.

Solution. Let $3x+4=t$. Then $3\, dx=dt$.

$$\int \csc (3x+4)\, dx=\frac{1}{3}\int \csc t\, dt=\frac{1}{3}\left[ln \left(\left|\tan \frac{t}{2}\right|\right)\right]+C,$$

$$=\frac{1}{3}\left[ln \left|\tan \left(\frac{3x+4}{2}\right)\right|\right]+C.$$

Example 17. *Evaluate* $\int \frac{dx}{\sqrt{(1+\cos 2x)}}$.

Solution. Now

$$\int \frac{dx}{\sqrt{(1+\cos 2x)}}=\int \frac{dx}{\sqrt{(1+2 \cos^2 x-1)}},$$

$$=\int \frac{dx}{\sqrt{2} . \cos x}$$

$$=\frac{1}{\sqrt{2}}\int \sec x\, dx,$$

$$=\frac{1}{\sqrt{2}}\,(ln \mid \sec x+\tan x \mid)+C.$$

EXERCISE 5 (*c*)

Evaluate each of the following :

1. $\int (10x+7)(5x^2+7x+3)\, dx.$

2. $\int (15x^2+14x)(5x^3+7x^2+3)\, dx.$

3. $\int x\, e^{x^2}\, dx.$

4. $\int \sec^2 x\, e^{\tan x}\, dx.$

5. $\int \frac{dx}{9x+1}.$

6. $\int \frac{\cos x-\sin x}{\cos x+\sin x}\, dx.$

7. $\int \frac{2x}{x^2+1}\, dx.$

8. $\int \frac{e^x}{1+e^x}\, dx.$

9. $\int \frac{dx}{(x+5)^2}.$

10. $\int \sec^2 (x-3)\, dx.$

11. $\int (3x-7)^8\, dx.$

12. $\int 9^{11x}\, dx.$

13. $\int \frac{x\, dx}{\sqrt{(7x^2+5)}}.$

14. $\int 3\sqrt{\cos x}\, \sin x\, dx.$

15. $\int 5 \sin x \sec^4 x\, dx.$

16. $\int \frac{\cos 2x+2\sin^2 x}{\cos^2 x}\, dx.$

17. $\int \frac{x^2\, dx}{1+x^6}.$

18. $\int \frac{e^x}{e^x+e^{-x}}\, dx.$

19. $\int \frac{\cos (5\, ln\, x-2)}{x}\, dx.$

20. $\int \frac{dx}{1+\cos x}.$

21. $\int \frac{\cot x\, dx}{ln \sin x}.$

22. $\int \frac{\cos x \sin x\, dx}{5\cos^2 x+3\sin^2 x}.$

23. $\int \tan (4x+11)\, dx.$

24. $\int (1-\sec x)^2\, dx.$

25. $\int \frac{5}{\sqrt{(1-\cos^2 x)}}\, dx.$

26. $\int \frac{7}{\sin x \cos x}\, dx.$

27. $\int \frac{\sin (x+2)}{\sin x}\, dx.$

28. $\int \frac{\cos (x+a)}{\cos x}\, dx.$

29. $\int \frac{1}{\sqrt{x} \sin \sqrt{x}}\, dx.$

30. $\int \frac{\sec (\tan^{-1} x)}{1+x^2}\, dx.$

31. $\int \frac{\cot (\sin^{-1} x)}{\sqrt{(1-x^2)}}\, dx.$

32. $\int \frac{\tan (ln\, x)}{x}\, dx.$

5·3·2. Integrals of the type

$$\int \frac{\mathbf{dx}}{\mathbf{x^2+a^2}},\quad \int \frac{\mathbf{dx}}{\mathbf{a^2-x^2}}$$

(A) Let $\quad I=\int \frac{dx}{x^2+a^2}$.

Put $x=a \tan\theta$. Then $dx=a \sec^2\theta\, d\theta$,

$$I=\int \frac{a \sec^2\theta\, d\theta}{a^2 \tan^2\theta+a^2}$$

$$=\frac{1}{a}\int d\theta.$$

$$=\frac{1}{a}\,.\,\theta+C$$

$$=\frac{1}{a}\tan^{-1}(x/a)+C.$$

Ramark. If we make the substitution $x=a \cot\theta$, the value of the integral can be seen to be $-(1/a)\cot^{-1}(x/a)+C$. The two values, however, are not different.

(B) Let $\quad I=\int \frac{dx}{x^2-a^2}$.

Writing $\quad \frac{1}{x^2-a^2}=\frac{1}{2a}\left[\frac{1}{x-a}-\frac{1}{x+a}\right]$,

we have

$$I=\int \frac{1}{2a}\left(\frac{1}{x-a}-\frac{1}{x+a}\right)dx$$

$$=\frac{1}{2a}\left[\ln|x-a|-\ln|x+a|\right]+C$$

$$=\frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right|+C.$$

Let $\quad I=\int \frac{dx}{a^2-x^2}$.

Writing $\frac{1}{a^2-x^2}=\frac{1}{2a}\left[\frac{1}{a+x}+\frac{1}{a-x}\right]$,

We have

$$I=\int \frac{1}{2a}\left(\frac{1}{a+x}+\frac{1}{a-x}\right)dx,$$

$$=\frac{1}{2a}\left(\ln|a+x|-\ln|a-x|\right)+C,$$

$$=\frac{1}{2a}\ln\left|\frac{a+x}{a-x}\right|+C.$$

Remark. Observe that the integrals in (B) and (C) express the same fact. The form in (B) is generally used when $x > a$, and that in (C) is used when $x < a$.

Integrals of the type

$$\int \frac{dx}{ax^2+bx+c}, \quad \int \frac{px+q}{ax^2+bx+c}\, dx.$$

(A) Let $\quad I = \int \frac{dx}{ax^2+bx+c}.$

We write $ax^2+bx+c = a\left(x^2+\frac{b}{a}x+\frac{c}{a}\right)$

$$= a\left\{\left(x+\frac{b}{2a}\right)^2 + \frac{4ac-b^2}{4a^2}\right\} \qquad \cdots(1)$$

If $4ac-a^2>0$, then $\frac{4ac-b^2}{4a^2} = m^2$ for some real number m, and therefore

$$I = \frac{1}{a}\int \frac{dx}{\left(x+\frac{b}{2a}\right)^2 + m^2} \qquad \cdots(2)$$

If $4\,ac-b^2<0$, then $\frac{4\,ac-b^2}{4a^2} = -m^2$, for some real number m, and therefore

$$I = \frac{1}{a}\int \frac{dx}{\left(x+\frac{b}{2a}\right)^2 - m^2} \qquad \cdots(3)$$

From (2) and (3) we find that in both the above cases, I is of the type discussed in 5·2·3 (A) and (B) above, and can, therefore, be integrated.

If $4\,ac-b^2=0$, then

$$I = \frac{1}{a}\int \frac{dx}{\left(x+\frac{b}{2a}\right)^2}$$

$$= -\frac{1}{a}\cdot\frac{1}{\left(x+\frac{b}{2a}\right)} + C.$$

(B) Let $\quad I = \int \frac{px+q}{ax^2+bx+c}\, dx$

Since the derivative of $ax^2+bx+c = 2ax+b$, we write

$$px+q \equiv l(2ax+b)+m, \qquad \cdots(1)$$

where l and m are to be determined.

Comparing the coefficients of x on both sides of (1), we have

$$p=2\,al. \quad ...(2)$$

Also, by equating the constant terms on both sides of (1), we have

$$q=bl+m \quad ...(3)$$

From (2) and (3) we have $l=p/(2a)$, $m=q-bp/(2a)$.

We can write

$$I=\int \frac{l(2ax+b)+m}{ax^2+bx+c}dx,$$

$$=l\int \frac{2ax+b}{ax^2+bx+c}\,dx+m\,J,$$

$$=l\,.\,ln\mid ax^2+bx+c\mid+m\,J,$$

where $$J=\int \frac{dx}{ax^2+bx+c}.$$

Since J can be evaluated as in (A) above, therefore I can be evaluated.

Example 18. *Evaluate* :

(*a*) $\int \frac{dx}{x^2+9}$ (*b*) $\int \frac{dx}{4x^2-25}$ (*c*) $\int \frac{dx}{16-9x^2}$.

Solution.

(*a*) The integrand is of the form $\frac{1}{x^2+a^2}$ with $a=3$.

$\therefore$ $$I=\int \frac{dx}{x^2+9}$$

$$=\frac{1}{3}\tan^{-1}(x/3)+C.$$

(*b*) $$I=\int \frac{dx}{4x^2-25}$$

$$=\frac{1}{2}\int \frac{2\,dx}{(2x)^2-5^2}$$

$$=\frac{1}{2\,.\,2\,.\,5}\,ln\left|\frac{2x-5}{2x+5}\right|+C$$

$$=\frac{1}{20}\,ln\left|\frac{2x-5}{2x+5}\right|+C.$$

(*c*) $$I=\int \frac{dx}{16-9x^2}$$

$$=\frac{1}{3}\int \frac{3\,dx}{4^2-(3x)^2}$$

$$=\frac{1}{3\,.\,2\,.\,4}\,ln\left|\frac{4+3x}{4-3x}\right|+C$$

$$=\frac{1}{24}\;ln\left|\frac{4+3x}{4-3x}\right|+C.$$

Example 19. *Evaluate :*

(*a*) $\int\frac{dx}{(2x^2+x+5)}$ (*b*) $\int\frac{(3x+4)}{2x^2+3x-4}\,dx.$

Solution.

(*a*) We can write

$$2x^2+x+5=2\left(x^2+\frac{1}{2}x+\frac{5}{2}\right)$$

$$=2\left\{\left(x+\frac{1}{4}\right)^2-\left(\frac{1}{4}\right)^2+\frac{5}{2}\right\}$$

$$=2\left\{\left(x+\frac{1}{4}\right)^2+\left(\frac{\sqrt{39}}{4}\right)^2\right\}$$

$$\therefore \qquad I=\int\frac{dx}{2x^2+x+5}$$

$$=\frac{1}{2}\int\frac{dx}{\left\{\left(x+\frac{1}{4}\right)^2+(\sqrt{39}/4)^2\right\}},$$

$$=\frac{1}{2}\cdot\frac{4}{\sqrt{(39)}}\tan^{-1}\left\{\left(x+\frac{1}{4}\right)\Big/\left(\sqrt{39}/4\right)\right\}+C,$$

$$=\frac{2}{\sqrt{39}}\tan^{-1}\{(4x+1)/\sqrt{39}\}+C.$$

(*b*) Since the derivative of $2x^2+3x-4$ is $4x+3$, we write

$$3x+4\equiv l(4x+3)+m.$$

Comparing coefficients of x and the constant terms on both sides of the above relation, we have

$$4l=3,\ 3l+m=4.$$

$$\therefore \qquad l=\frac{3}{4},\ m=4-3l=4-\frac{9}{4}=\frac{7}{4}.$$

$$\therefore \quad \frac{3x+4}{2x^2+3x-4}=\frac{\frac{3}{4}(4x+3)+\frac{7}{4}}{2x^2+3x-4}$$

$$I=\int\frac{3x+4}{2x^2+3x-4}\,dx$$

$$=\frac{3}{4}\int\frac{4x+3}{2x^2+3x-4}\,dx+\frac{7}{4}\int\frac{dx}{2x^2+3x-4}$$

$$=\frac{3}{4}I_1+\frac{7}{4}I_2, \qquad \ldots(1)$$

where $$I_1=\int\frac{4x+3}{2x^2+3x-4}\,dx$$

$$=ln\mid 2x^2+3x-4\mid+C_1, \qquad \ldots(2)$$

and $$I_2=\int\frac{dx}{2x^2+3x-4} \qquad \ldots(3)$$

Now $2x^2+3x-4=2\left(x^2+\frac{3}{2}x-2\right)$

$$=2\left\{\left(x+\frac{3}{4}\right)^2-\left(\frac{5}{2}\right)^2\right\}.$$

$\therefore$ $$I_2=\frac{1}{2}\int\frac{dx}{\left(x+\frac{3}{4}\right)^2-\left(\frac{5}{4}\right)^2}$$

$$=\frac{1}{5}\,ln\left|\frac{x+\frac{3}{4}-\frac{5}{4}}{x+\frac{3}{4}+\frac{5}{4}}\right|+C_2$$

$$=\frac{1}{5}\,ln\left|\frac{2x-1}{2x+4}\right|+C_2 \qquad \ldots(4)$$

From (1), (3) and (4) we find that

$$I=\frac{3}{4}I_1+\frac{7}{4}I_2$$

$$=\frac{3}{4}\,ln\mid 2x^2+3x-4\mid+\frac{7}{20}\,ln\left|\frac{2x-1}{2x+4}\right|+C,$$

where we have written C for $\frac{3}{4}C_1+\frac{7}{4}C_2$.

EXERCISE 5 (*d*)

Evaluate :

1. $\int\frac{dx}{x^2+4}$.
2. $\int\frac{dx}{9-x^2}$.
3. $\int\frac{dx}{4x^2+1}$.
4. $\int\frac{dx}{x^2-25}$.
5. $\int\frac{dx}{4x^2-4x+2}$.
6. $\int\frac{dx}{x^2+2x+3}$

7. $\int \frac{dx}{2x^2+x-1}$. 8. $\int \frac{dx}{x^2-3x+2}$.

9. $\int \frac{3x+1}{2x^2+x+1}\,dx$. 10. $\int \frac{3x}{x^2-x-2}\,dx$.

11. $\int \frac{5x-2}{1+2x+3x^2}\,dx$. 12. $\int \frac{x}{x^2+2x+3}\,dx$.

5·3·3. Integrals of the type

$$\int \frac{dx}{\sqrt{(a^2-x^2)}},\quad \int \frac{dx}{\sqrt{(x^2\pm a^2)}}.$$

(A) Let $\mathrm{I}=\int \frac{dx}{\sqrt{(a^2-x^2)}}$.

Put $x=a\sin\theta$. Then $dx=a\cos\theta\,d\theta$

$$\begin{aligned}\mathrm{I}&=\int \frac{a\cos\theta\,d\theta}{\sqrt{(a^2-a^2\sin^2\theta)}}\\ &=\int d\theta\\ &=\theta+\mathrm{C}\\ &=\sin^{-1}(x/a)+\mathrm{C}.\end{aligned}$$

Remark. If we make the substitution $x=a\cos\theta$, the value of the integral can be seen to be $-\cos^{-1}(x/a)+\mathrm{C}$. Of course, because of the fact that $\sin^{-1}x+\cos^{-1}x=\pi/2$, the two values are not different, in the sense that each expression represents the same family of primitives.

(B) Let $\mathrm{I}=\int \frac{dx}{\sqrt{x^2+a^2}}$.

Put $x=a\tan\theta$. Then $dx=a\sec^2\theta\,d\theta$.

$$\begin{aligned}\mathrm{I}&=\int \frac{a\sec^2\theta\,d\theta}{\sqrt{(a^2\tan^2\theta)+a^2}}\\ &=\int \sec\theta\,d\theta\\ &=ln\,|\sec\theta+\tan\theta\,|+\mathrm{C}\\ &=ln\left(\left|\frac{\sqrt{(x^2+a^2)}+x}{a}\right|\right)+\mathrm{C}\\ &=ln\,|\sqrt{(x^2+a^2)}+x\,|+\mathrm{C}',\end{aligned}$$

where we have written C' for $\mathrm{C}-ln\,|a|$.

(C) Let $\mathrm{I}=\int \frac{dx}{\sqrt{(x^2-a^2)}}$.

Put $x=a \sec \theta$. Then $dx=a \sec \theta \tan \theta \, d\theta$

$$\therefore \quad I=\int \frac{a \sec \theta \tan \theta}{\sqrt{(a^2 \sec^2 \theta-a^2)}} \, d\theta$$

$$=\int \sec \theta \, d\theta$$

$$=ln \mid \sec \theta+\tan \theta \mid +C$$

$$=ln\left|\frac{x+\sqrt{(x^2-a^2)}}{a}\right|+C$$

$$=ln \mid \sqrt{(x^2-a^2)}+x \mid +C',$$

where we have written C′ for $C-ln \mid a \mid$.

Example 20. *Evaluate* :

(*a*) $\displaystyle\int \frac{dx}{\sqrt{(16-x^2)}}$ (*b*) $\displaystyle\int \frac{dx}{\sqrt{(4x^2+9)}}$

(*c*) $\displaystyle\int \frac{\cos x \, dx}{\sqrt{(16 \sin^2 x-9}}$.

Solution.

(*a*) The integrand is of the form

$$\frac{1}{\sqrt{(a^2-x^2)}} \text{ with } a=4.$$

$$\therefore \quad \int \frac{dx}{\sqrt{(16-x^2)}}=\sin^{-1}(x/4)+C.$$

(*b*)
$$I=\int \frac{dx}{\sqrt{(4x^2+9)}}$$

$$=\frac{1}{2}\int \frac{2 \, dx}{\sqrt{((2x)^2+3)^2}},$$

$$=\frac{1}{2} ln \mid \sqrt{\{(2x)^2+3^2\}}+2x \mid +C,$$

$$=\frac{1}{2} ln \mid \sqrt{(4x^2+9)}+2x \mid +C.$$

(*c*)
$$I=\int \frac{\cos x \, dx}{\sqrt{(16 \sin^2 x-9)}}.$$

Put $\sin x=t$. Then $\cos x \, dx=dt$.

$$I=\int \frac{dt}{\sqrt{(16t^2-9)}}$$

$$=\frac{1}{4}\int \frac{4 \, dt}{\sqrt{[(4t)^2-3^2]}}$$

$$=\frac{1}{4}\, ln \mid \sqrt{[(4t)^2-3^2]}+4t \mid +C$$

$$=\frac{1}{4}\, ln \mid \sqrt{(16 \sin^2 x-9)}+4 \sin x \mid +C.$$

5·3·4. Integrals of the type

$$\int \frac{dx}{\sqrt{(ax^2+bx+c)}} \quad \int \frac{px+q}{\sqrt{(ax^2+bx+c)}}\, dx.$$

(A) Let $\quad I=\int \frac{dx}{\sqrt{(ax^2+bx+c)}}.$

The method consists in reducing the integrand to one of the standard forms discussed above.

We write $ax^2+bx+c=a\left(x^2+\frac{b}{a}x+\frac{c}{a}\right)$

$$=\left\{\left(x+\frac{b}{2a}\right)^2+\frac{4ac-b^2}{4a^2}\right\}.$$

Four different cases arise ;

(*i*) $a>0,\ 4ac-b^2>0.$

We can write $\frac{4ac-b^2}{4a^2}=k^2$, for some real number k.

Then $\quad I=\frac{1}{\sqrt{a}}\int \frac{dx}{\sqrt{\{[x+c(b/2a)]^2+k^2\}}},$

(*ii*) $a>0,\ 4ac-b^2<0.$

We can write $\frac{4ac-b^2}{4a^2}=-k^2$, for some real number k.

Then $\quad I=\frac{1}{\sqrt{a}}\int \frac{dx}{\sqrt{[\{x+b/2a)\}^2-k^2]}}.$

(*iii*) $a<0,\ 4ac-b^2>0.$ In this case $ax^2+bx+c<0$ for all real values of x, and hence the integrand does not exist.

(*iv*) $a<0,\ 4ac-b^2<0.$

We can write $a=-h,\ h>0, \frac{4ac-b^2}{4a^2}=-k^2$ for some real number k.

Then $\quad I=\frac{1}{\sqrt{h}}\int \frac{dx}{\sqrt{[k^2-\{x-(b/2h)\}^2]}}$

Thus we find that in all the cases (where the integrand has a meaning) I can be reduced to one of the standard forms discussed above.

(B) Let $$I=\int \frac{px+q}{\sqrt{(ax^2+bx+c)}}\,dx$$

Observe that the derivative of ax^2+bx+c is $2ax+b$. We therefore write

$$(px+q)\equiv l(2ax+b)+m.$$

Comparing coefficients of x on both side of the above identity, we have

$$p=2al. \qquad ...(i)$$

Also, comparing the constant terms on both sides, we have

$$q=bl+m. \qquad ...(ii)$$

From (i) and (ii), $l=p/(2a)$, $m=q-bp/(2a)$ ...(iii)

With these values of l and m, we can write

$$I=\int \frac{l(2ax+b)+m}{\sqrt{(ax^2+bx+c)}}\,dx$$

$$=l\int \frac{2ax+b}{\sqrt{(ax^2+bx+c)}}\,dx+m\int \frac{dx}{\sqrt{(ax^2+bx+c)}}$$

$$=2l\sqrt{(ax^2+bx+c)}+m\,J,$$

where $$J=\int \frac{dx}{\sqrt{(ax^2+bx+c)}}.$$

The method of integrating J has been discussed in (A) above, and therefore we can evaluate I.

Remark. The substitution $t=x+(b/2a)$ can also be used to reduce the integrals discussed above to a standard form.

The following examples will illustrate how integrals of the above type can be evaluated.

Example 21. *Evaluate* :

(a) $\int \frac{dx}{\sqrt{(x^2+2x+5)}}$ (b) $\int \frac{dx}{\sqrt{(3x-x^2-2)}}$.

Solution.

(a) $$\int \frac{dx}{\sqrt{(x^2+2x+5)}}=\int \frac{dx}{\sqrt{\{(x+1)^2+4\}}}$$

$$=\ln\left|\sqrt{\{(x+1)^2+4\}}+(x+1)\right|+C,$$

$$=\ln\left|\sqrt{(x^2+2x+5)}+x+1\right|+C.$$

(b) We write

$$3x-x^2-2=-2-(x^2-3x)$$

$$=-2-\left[\left(x-\frac{3}{2}\right)^2-\frac{9}{4}\right]$$

$$=\left(\frac{9}{4}-2\right)-\left(x-\frac{3}{2}\right)^2$$

$$=\left(\frac{1}{2}\right)^2-\left(x-\frac{3}{2}\right)^2.$$

$$\therefore\quad \int\frac{dx}{\sqrt{(3x-x^2-2)}}=\int\frac{dx}{\sqrt{\left[\left(\frac{1}{2}\right)^2-\left(x-\frac{3}{2}\right)^2\right]}}$$

$$=\sin^{-1}\left[2\left(x-\frac{3}{2}\right)\right]+C$$

$$=\sin^{-1}(2x-3)+C.$$

Example 22. *Evaluate* :

(*a*) $\int\frac{x+2}{\sqrt{(x^2-2x+4)}}\,dx$ (*b*) $\int\frac{2x+3}{\sqrt{(3+4x-4x^2)}}\,dx.$

Solution.

(*a*) Since the derivative of $x^2-2x+4=2x-2$, we write the numerator $x+2$ as $\frac{1}{2}(2x-2)+3$. Then

$$I=\int\frac{(x+2)}{\sqrt{(x^2-2x+4)}}\,dx$$

$$=\int\frac{\frac{1}{2}(2x-2)+3}{\sqrt{(x^2-2x+4)}}\,dx$$

$$=\frac{1}{2}\int\frac{2x-2}{\sqrt{(x^2-2x+4)}}\,dx+3\int\frac{dx}{\sqrt{(x^2-2x+4)}}$$

$$=(x^2-2x+4)^{\frac{1}{2}}+3\int\frac{dx}{\sqrt{[(x-1)^2+(\sqrt{3})^2]}}$$

$$=(x^2-2x+4)^{\frac{1}{2}}+3\ln\left|\sqrt{\{(x-1)^2+(\sqrt{3})^2\}}+(x-1)\right|+C$$

$$=(x^2-2x+4)^{\frac{1}{2}}+3\ln\left|\sqrt{(x^2-2x+4)}+x-1\right|+C.$$

(*b*) The expression under the radical sign in the denominator can be written as

$$3-(4x^2-4x)=4-(2x-1)^2$$

We therefore try the substitution $2x-1=t$, *i.e.*, $x=\frac{1}{2}(t+1)$. Then, $dx=\frac{1}{2}\,dt$.

$$I=\int\frac{2x+3}{\sqrt{(3+4x-4x^2)}}\,dx,$$

$$=\int\frac{t+4}{\sqrt{(4-t^2)}}\cdot\frac{1}{2}\,dt,$$

$$=\frac{1}{2}\int\frac{t}{\sqrt{(4-t^2)}}\,dt+2\int\frac{dt}{\sqrt{(4-t^2)}},$$

$$=-\frac{1}{2}(4-t^2)^{\frac{1}{2}}+2\sin^{-1}(t/2)+C,$$

$$=-\frac{1}{2}(3+4x-4x^2)^{\frac{1}{2}}+2\sin^{-1}\left(x-\frac{1}{2}\right)+C.$$

Remark. In the above example, we used slightly different methods for evaluating the two integrals, simply for the sake of illustration. Both the integrals could have been evaluated by either of the two methods.

EXERCISE 5 (*e*)

Evaluate :

1. $\int\frac{dx}{\sqrt{(9-x^2)}}$.
2. $\int\frac{dx}{\sqrt{(16+x^2)}}$.
3. $\int\frac{dx}{\sqrt{(1-4x^2)}}$.
4. $\int\frac{dx}{\sqrt{(1+9x^2)}}$.
5. $\int\frac{dx}{\sqrt{(x^2-16)}}$.
6. $\int\frac{dx}{\sqrt{(4x^2-9)}}$.
7. $\int\frac{dx}{\sqrt{(9-16x^2)}}$.
8. $\int\frac{e^x}{\sqrt{25x^2+16}}$.
9. $\int\frac{x\,dx}{\sqrt{(1-x^4)}}$.
10. $\int\frac{e^x}{\sqrt{(e^{2x}+1)}}\,dx$.
11. $\int\frac{x^2\,dx}{\sqrt{(x^6+1)}}$.
12. $\int\frac{dx}{\sqrt{(x^2+4x)}}$.
13. $\int\frac{dx}{\sqrt{(2x^2+3x+4)}}$.
14. $\int\frac{dx}{\sqrt{(x^2+x+1)}}$.
15. $\int\frac{dx}{\sqrt{(3x^2-x+2)}}$.
16. $\int\frac{x\,dx}{\sqrt{(8+x-x^2)}}$.
17. $\int\frac{x+1}{\sqrt{(x^2-x+1)}}\,dx$.
18. $\int\frac{2x+5}{\sqrt{(x^2+3x+1)}}\,dx$.

5 4. INTEGRATION BY PARTS

The method of integration by parts is a method of determining primitives of functions expressible as the product of two functions whose primitives are known. It consists essentially in expressing the integral of a product of two functions in terms of another integral whose evaluation may be simpler.

Theorem 5·8. *Let $f(x)$ and $g(x)$ be two functions such that $\int g(x)\,dx$ and $\int\left[f'(x)\int g(x)\,dx\right]dx$ exist. Then*

$$\int f(x)g(x)\,dx=f(x)\int g(x)\,dx-\int\left[f'(x)\int g(x)\,dx\right]dx.$$

Proof. We have,

$$\frac{d}{dx}\left[f(x)\int g(x)\,dx\right]=f(x)\cdot\frac{d}{dx}\left[\int g(x)\,dx\right]+f'(x)\int g(x)\,dx,$$

$$=f(x)g(x)+f'(x)\int g(x)\,dx.$$

$$\therefore\ f(x)g(x)=\frac{d}{dx}\left[f(x)\cdot\int g(x)\,dx\right]-f'(x)\int g(x)\,dx.$$

Integrating both sides, we get

$$\int f(x)\,g(x)\,dx=f(x)\int g(x)\,dx-\int f'(x)\left[\int g(x)\,dx\right]dx.$$

This proves the theorem.

Remarks 1. In the above theorem, $f(x)$ is referred to as the first function and $g(x)$ is referred to as the second function. Naturally, the theorem will be useful only when it is easier to evaluate the integrals $\int g(x)dx$ and $\int f'(x)[\int g(x)dx]dx$ than to evaluate the given integral. The choice of the first function and the second function is therefore not arbitrary. We would like to label that function as the first function whose derivative is simpler in form and that function as the second function whose integral is easy to evaluate.

2. Expressed differently, the theorem says :

The integral of the product of two functions = first function × integral of second − integral of (derivative of first × integral of second).

3. By writing $u=f(x)$ and $v=\int g(x)dx$, another simple way to remember the result is

$$\int uv'\,dx=uv-\int u'v\,dx,\text{ where } u'=f'(x) \text{ etc.}$$

We shall illustrate the method by means of some examples. It may be noted that in some cases more than one application of the theorem may be necessary.

Example 23. Evaluate $\int x\cos x\,dx$.

Solution. Take x as the first function and $\cos x$ as the second. Now integrating by parts, we get

$$\int x \cos x \, dx = x\int \cos x \, dx - \int\left[\frac{d}{dx}(x)\int \cos x \, dx\right]dx,$$

$$= x \sin x - \int 1.\left[\int \cos x \, dx\right]dx,$$

$$= x \sin x - \int \sin x \, dx,$$

$$= x \sin x + \cos x + C.$$

Remarks. Strictly speaking, we ought to add the constant of integration each time that we write the primitive. However, it can be easily seen that it suffices to add the constant of integration at the end.

Example 24. *Evaluate* $\int \sin^{-1} x \, dx$.

Solution. Here the integrand is *not* actually the product of two functions, but we shall still apply the method of integration by parts by writing the integrand as

$(\sin^{-1}x).\ 1$, and taking first function $= \sin^{-1}x$

and second function $= 1$.

Then $I = \int (\sin^{-1}x)\ 1\ dx,$

$$= (\sin^{-1}x).\ x - \int \frac{1}{\sqrt{(1-x^2)}}.\ x \, dx,$$

$$= x \sin^{-1}x + \sqrt{(1-x^2)} + C.$$

Example 25. *Evaluate* $\int \ln x \, dx$.

Solution. Now $\int \ln x dx = \int \ln x.\ 1\ dx.$

Integrating by parts,

$$\int \ln x.\ 1\ dx = \ln x \int 1 .\ dx - \int\left[\frac{1}{x}\int 1.\ dx\right]dx,$$

$$= (\ln x).\ x - \int \frac{1}{x}.\ x \, dx,$$

$$= (\ln x).\ x - \int 1.\ dx,$$

$$= x \ln x - x + C.$$

Example 26. *Evaluate* $\int x^2 e^x dx$.

Solution. Integrating by parts,

$$\int x^2 e^x\,dx = x^2 \int e^x\,dx - \int\left[\frac{d}{dx}(x^2)\int e^x\,dx\right]dx,$$

$$= x^2 e^x - \int (2x \,.\, e^x)\,dx.$$

$$= x^2 e^x - 2\int x e^x\,dx. \qquad \ldots(i)$$

Again, integrating by parts,

$$\int x e^x dx = x\int e^x dx - \int\left[\frac{d}{dx}(x)\int e^x dx\right]dx,$$

$$= xe^x - \int 1 \,.\, e^x dx = xe^x - e^x. \qquad \ldots(ii)$$

From (*i*) and (*ii*), we get

$$\int x^2 e^x dx = x^2 e^x - 2\,[xe^x - e^x] + C = e^x\,(x^2 - 2x + 2) + C.$$

Example 27. *Evaluate* $\int \frac{x \sin^{-1} x}{\sqrt{(1-x^2)}}\,dx.$

Solution. Let us put $\sin^{-1} x = t$.

Then $\frac{1}{\sqrt{(1-x^2)}}\,dx = dt.$

$$\therefore \qquad I = \int \frac{x \sin^{-1} x}{\sqrt{(1-x^2)}}\,dx = \int t \sin t\,dt. \qquad \ldots(1)$$

Let us now integrate $t \sin t$ by parts, taking first function $= t$, and second function $= \sin t$.

Then $\qquad I = \int t \sin t\,dt,$

$$= t(-\cos t) - \int 1 \,.\, (-\cos t)\,dt,$$

$$= -t \cos t + \int \cos t\,dt,$$

$$= -t \cos t + \sin t + C,$$

$$= -(\sin^{-1} x)\sqrt{(1-x^2)} + x + C.$$

Thus

$$\int \frac{x \sin^{-1} x}{\sqrt{(1-x^2)}}\,dx = -\sqrt{(1-x^2)}\,\sin^{-1} x + x + C.$$

Example 28. *Evaluate* $\int \frac{xe^x}{(1+x)^2}\,dx.$

Solution. $$I=\int \frac{xe^x}{(1+x)^2}\,dx=\int \frac{(1+x)-1}{(1+x)^2}\,e^x\,dx,$$
$$=\int e^x\left\{\frac{1}{1+x}-\frac{1}{(1+x)^2}\right\}dx,$$
$$=\int \frac{e^x}{1+x}\,dx-\int \frac{e^x}{(1+x)^2}\,dx. \qquad ...(1)$$

Let us integrate $\frac{e^x}{1+x}$ by parts taking first function $=\frac{1}{1+x}$, second function $=e^x$. Then
$$\int \frac{e^x}{1+x}\,dx=\frac{1}{1+x}\cdot e^x-\int -\frac{1}{(1+x)^2}\cdot e^x\,dx+C,$$
$$=\frac{e^x}{1+x}+\int \frac{e^x}{(1+x)^2}\,dx+C,$$
or $$\int \frac{e^x}{1+x}\,dx-\int \frac{e^x}{(1+x)^2}\,dx=\frac{e^x}{1+x}+C,$$
i.e., $$I=\frac{e^x}{1+x}+C.$$

Remark. In the above example we expressed the integrand in the form $e^x(f(x)+f'(x))$, where $f(x)=\frac{1}{1+x}$ and integrated $e^xf(x)$. It was found that the final integral came out to be simply $\int e^xf(x)+C$. This is because of the following fact :

Since $D(e^xf(x))=(De^x)f(x)+e^xDf(x)$,
$$=e^xf(x)+e^xf'(x)=e^x(f(x)+f'(x)),$$
therefore,
$$\int e^x(f(x)+f'(x))dx=e^xf(x)+C.$$

You will find the above observation useful.

5·4·1. Integration of e^{ax} cos bx and e^{ax} sin bx

(*i*) $$\int e^{ax}\cos bx\,dx=\frac{e^{ax}}{\sqrt{(a^2+b^2)}}\cos\left(bx-\tan^{-1}\frac{b}{a}\right)+C,$$

(*ii*) $$\int e^{ax}\sin bx\,dx=\frac{e^{ax}}{\sqrt{(a^2+b^2)}}\sin\left(bx-\tan^{-1}\frac{b}{a}\right)+C.$$

Integrating by parts,
$$\int e^{ax}\cos bx\,dx=e^{ax}\int \cos bx\,dx-\int\left[\frac{d}{dx}(e^{ax})\int \cos bx\,dx\right]dx,$$
$$=\frac{1}{b}e^{ax}\sin bx-\frac{a}{b}\int e^{ax}\sin bx\,dx. \qquad ...(i)$$

Also, $\int e^{ax} \sin bx\, dx$

$$= e^{ax}\int \sin bx\, dx - \int\left[\frac{d}{dx}(e^{ax})\int \sin bx\, dx\right]dx,$$

$$= -e^{ax}\frac{\cos bx}{b} + \frac{a}{b}\int e^{ax}\cos bx\, dx. \qquad \text{...}(ii)$$

Let $u=\int e^{ax}\cos bx\, dx$, and $v=\int e^{ax}\sin bx\, dx$.

Then (i) and (ii) can be written respectively as

$$u=\frac{1}{b}e^{ax}\sin bx-\frac{a}{b}v, \qquad \text{...}(iii)$$

and
$$v=-\frac{1}{b}e^{ax}\cos bx+\frac{a}{b}u. \qquad \text{...}(iv)$$

Substituting the value of v from (iv) in (iii), we get

$$u=\frac{1}{b}\cdot e^{ax}\sin bx-\frac{a}{b}\left(-\frac{1}{b}e^{ax}\cos bx + \frac{a}{b}u\right).$$

This gives,

$$u=\frac{e^{ax}}{a^2+b^2}(a\cos bx+b\sin bx).$$

Similarly, we get

$$v=\frac{e^{ax}}{a^2+b^2}(a\sin bx-b\cos bx),$$

Now, if we put $a=r\cos\theta$, $b=r\sin\theta$,

then, $r=\sqrt{(a^2+b^2)}$, $\theta=\tan^{-1}\frac{b}{a}$,

Also, then

$$u=\frac{e^{ax}}{a^2+b^2}\cdot r(\cos\theta\cos bx+\sin\theta\sin bx),$$

$$=\frac{e^{ax}}{\sqrt{(a^2+b^2)}}\cos\left(bx-\tan^{-1}\frac{b}{a}\right). \qquad \text{...}(v)$$

Similarly,

$$v=\frac{e^{ax}}{\sqrt{(a^2+b^2)}}\sin\left(bx-\tan^{-1}\frac{b}{a}\right). \qquad \text{...}(vi)$$

Since any two primitives of a function differ by a constant, therefore to get the general forms of the primitives, we add the constants of integration to the values of u and v obtained above.

We thus have

$$\int e^{ax}\cos bx\,dx=\frac{e^{ax}}{\sqrt{(a^2+b^2)}}\cos\,(bx-\tan^{-1} b/a)+C_1,$$

$$\int e^{ax}\sin bx\,dx=\frac{e^{ax}}{\sqrt{(a^2+b^2)}}\sin\,(bx-\tan^{-1} b/a)+C_2$$

Example 29. *Evaluate* $\int e^{2x}\sin 3x\,dx$.

Solution. Here $a=2$, $b=3$.

$$\therefore \int e^{2x}\sin 3x\,dx=\frac{e^{2x}}{\sqrt{(4+9)}}\left[\sin\left(3x-\tan^{-1}\frac{3}{2}\right)\right]+C,$$

$$=\frac{e^{2x}}{\sqrt{13}}\sin\left(3x-\tan^{-1}\frac{3}{2}\right)+C.$$

Example 30. *Evaluate* $\int e^{x}\cos^2 x\,dx$.

Solution. Now

$$\int e^{x}\cos^2 x\,dx=\int e^{x}.\frac{\cos 2x+1}{2}\,dx.$$

$$=\frac{1}{2}\int e^{x}(1+\cos 2x)\,dx,$$

$$=\frac{1}{2}\int e^{x}dx+\frac{1}{2}\int e^{x}\cos 2x\,dx.$$

$$=\frac{1}{2}e^{x}+\frac{1}{2}\left[\frac{e^{x}}{\sqrt{(1+4)}}\cos\,(2x-\tan^{-1}2)\right]+C,$$

$$=\frac{1}{2}e^{x}\left[1+\frac{1}{\sqrt{5}}\cos\,(2x-\tan^{-1}2)\right]+C.$$

EXERCISE 5 (*f*)

Evaluate each of the following :

1. $\int x\sin x\,dx.$

2. $\int x\,ln\,x\,dx.$ (*A.I.S.S.C.E., 1987*)

3. $\int \tan^{-1}x\,dx.$ (*A.I.S.S.C.E., 1985*)

4. $\int x^2\cos x\,dx.$

5. $\int x\,e^{2x}\,dx.$

6. $\int x^2 e^{-x}\,dx.$

7. $\int x \sin^{-1}x \, dx.$

8. $\int x^2 \sin x \cos x \, dx.$

9. $\int x \, ln \, (1+x) \, dx.$

10. $\int \frac{x}{\cos^2 x} dx.$

11. $\int \frac{x-\sin x}{1-\cos x} dx.$

1 . $\int (x^2+3x+4) \sin x \, dx.$

(*A.I.S.S.C.E., 1989*)

13. $\int \frac{1}{x} \, ln \, (ln \, (x)) \, dx.$

14. $\int x^2 \, ln \, x \, dx.$

15. $\int (ln \, x)^2 \, dx.$

16. $\int x \, (ln \, x)^2 \, dx.$

(*A.I.S.S.C.E., 1988*)

17. $\int x \tan^{-1}x \, dx.$

18. $\int \cos^{-1}x \, dx.$

19. $\int \tan^{-1} \sqrt{x} \, dx.$

20. $\int \frac{\sin^{-1}\sqrt{x}}{\sqrt{x}} dx.$

21. $\int \frac{\sin^{-1}x}{(1-x^2)^{3/2}} dx.$

22. $\int \frac{x^2 \tan^{-1}x}{1+x^2} dx.$

(*A.I.S.S.C.E., 1987*)

23. $\int \frac{x \tan^{-1}x}{(1+x^2)^{3/2}} dx.$

24. $\int \frac{e^x (x^2+1)}{(x+1)^2} dx.$

25. $\int e^x (\sin x + \cos x) \, dx.$

26. $\int e^x (\tan x - ln \mid \cos x \mid) dx.$

27. $\int \frac{e^x (1+\sin x)}{1+\cos x} dx.$

28. $\int e^x (\cot x + ln \mid \sin x \mid) dx.$

29. $\int \frac{e^x (1-\sin x)}{1-\cos x} dx.$

30. $\int e^x \left(\frac{1}{x} + ln \mid x \mid \right) dx.$

31. $\int e^{5x} \cos 6x \, dx.$

32. $\int e^x \sin 5x \, dx.$

33. $\int e^x \sin^2 x \, dx.$

34. $\int e^x \sin x \cos x \, dx.$

35. $\int \sin (ln \, x) \, dx.$

5·4·2. Integrals of the type $\int \sqrt{(a^2-x^2)} \, dx, \int \sqrt{(x^2 \pm a^2)} \, dx$

(*a*) Let $I = \int \sqrt{(a^2-x^2)} \, dx.$

Integrating by parts, taking first function $=\sqrt{(a^2-x^2)}$ and second function $=1$, we have

$$I=\int\sqrt{(a^2-x^2)}.\ 1\ dx$$

$$=\sqrt{(a^2-x^2)}.x-\int\left[\frac{1}{2}.\ \frac{(-2x)}{\sqrt{(a^2-x^2)}}\right].\ x\ dx$$

$$=x\ \sqrt{(a^2-x^2)}-\int\frac{-x^2}{\sqrt{(a^2-x^2)}}\ dx,$$

$$=x\sqrt{(a^2-x^2)}-\int\frac{(a^2-x^2)-a^2}{\sqrt{(a^2-x^2)}}\ dx,$$

$$=x\ \sqrt{(a^2-x^2)}-\int\sqrt{(a^2-x^2)}dx+a^2\int\frac{dx}{\sqrt{(a^2-x^2)}},$$

$$=x\ \sqrt{(a^2-x^2)}-I+a^2\sin^{-1}(x/a)+C_1,$$

or $\quad 2I=x\sqrt{(a^2-x^2)}+a^2\sin^{-1}(x/a)+C_1,$

or $\quad I=\frac{1}{2}x\ \sqrt{(a^2-x^2)}+\frac{1}{2}a^2\sin^{-1}x+C$, where we have written C for $\frac{1}{2}C_1$.

Aliter. We could obtain the above result by the substitution $x=a\sin\theta$ as well.

Put $x=a\sin\theta$, so that $dx=a\cos\theta\ d\theta$.

$$I=\int\sqrt{(a^2-a^2\sin^2\theta)}.\ a\cos\theta\ d\theta,$$

$$=a^2\int\cos^2\theta\ d\theta,$$

$$=a^2\int\frac{1}{2}(\cos 2\theta+1)\ d\theta,$$

$$=\frac{1}{2}\ a^2\left(\frac{1}{2}\sin 2\theta+\theta\right)+C,$$

$$=\frac{1}{2}a^2\left\{\frac{x}{a}\left[\sqrt{\left(1-\left(\frac{x}{a}\right)^2\right)}\right]+\sin^{-1}\frac{x}{a}\right\}+C,$$

$$=\frac{1}{2}x\ \sqrt{(a^2-x^2)}+\frac{1}{2}a^2\sin^{-1}(x/a)+C.$$

(*b*) Let $\quad I=\int\sqrt{(x^2-a^2)}\ dx$

Integrating by parts, taking first function $=\sqrt{(x^2-a^2)}$, and second function $=1$, we have

$$I=\sqrt{(x^2-a^2)}.\ x-\int\frac{1}{2}.\frac{2x}{\sqrt{(x^2-a^2)}}.\ x\ dx$$

$$=x\sqrt{(x^2-a^2)}-\int \frac{(x^2-a^2)+a^2}{\sqrt{(x^2-a^2)}}\,dx,$$

$$=x\sqrt{(x^2-a^2)}-\int \sqrt{(x^2-a^2)}\,dx-a^2\int \frac{dx}{\sqrt{(x^2-a^2)}},$$

$$=x\sqrt{(x^2-a^2)}-\mathrm{I}-a^2\,ln\,|\sqrt{(x^2-a^2)}+x\,|,$$

or $$2\mathrm{I}=x\sqrt{(x^2-a^2)}-a^2\,ln\,|\sqrt{(x^2-a^2)}+x\,|+\mathrm{C},$$

or $$\mathrm{I}=\tfrac{1}{2}x\sqrt{(x^2-a^2)}-\tfrac{1}{2}a^2\,ln\,|\sqrt{(x^2-a^2)}+x\,|+\mathrm{C}'$$

where we have written C' for $\frac{1}{2}\mathrm{C}$.

Aliter. Put $x=a\sec\theta$. Then $dx=a\sec\theta\tan\theta\,d\theta$

$$\mathrm{I}=\int \sqrt{(a^2\sec^2\theta-a^2)}\,.\,a\sec\theta\tan\theta\,d\theta,$$

$$=a^2\int \sec\theta\tan^2\theta\,d\theta,$$

$$=a^2\int \tan\theta\,(\sec\theta\tan\theta)\,d\theta. \qquad ...(1)$$

Integrate by parts, taking first function$=\tan\theta$, and second function$=\sec\theta\tan\theta$.

Then

$$\mathrm{I}=a^2[\tan\theta\sec\theta-\int \sec^2\theta\sec\theta\,d\theta],$$

$$=a^2[\tan\theta\sec\theta-\int (1+\tan^2\theta)\sec\theta\,d\theta],$$

$$=a^2[\tan\theta\sec\theta-\int \sec\theta\,d\theta-\mathrm{I}/a^2],$$

$$=a^2[\tan\theta\sec\theta-ln\,|\sec\theta+\tan\theta\,|\,]-\mathrm{I}+\mathrm{C},$$

or $$2\mathrm{I}=a^2\tan\theta\sec\theta-a^2\,ln\,|\sec\theta+\tan\theta\,|+\mathrm{C},$$

or $$\mathrm{I}=\tfrac{1}{2}x\sqrt{(x^2-a^2)}-\tfrac{1}{2}a^2\,ln\left|\frac{x+\sqrt{(x^2-a^2)}}{a}\right|+\mathrm{C},$$

$$=\tfrac{1}{2}x\sqrt{(x^2-a^2)}-\tfrac{1}{2}a^2\,ln\,|\,x+\sqrt{(x^2-a^2)}\,|+C',$$

where we have written C' for the constant $\frac{1}{2}a^2\,ln\,|\,a\,|+\mathrm{C}$.

(c) Let $\mathrm{I}=\int\sqrt{(x^2+a^2)}\,dx.$

Integrating by parts, taking first function$=\sqrt{(x^2+a^2)}$ and second function$=1$, we have

$$\mathrm{I}=\sqrt{(x^2+a^2)}\,.\,x-\int .\,\tfrac{1}{2}\,\frac{2x}{\sqrt{(x^2+a^2)}}\,.\,x\,dx,$$

$$=x\sqrt{(x^2+a^2)}-\int \frac{(x^2+a^2)-a^2}{\sqrt{(x^2+a^2)}}\,dx,$$

$$=x\sqrt{(x^2+a^2)}-\int \sqrt{(x^2+a^2)}\,dx+a^2\int \frac{dx}{\sqrt{(x^2+a^2)}},$$

$$=x\sqrt{(x^2+a^2)}-I+a^2\ ln\ |\ \sqrt{(x^2+a^2)}+x\ |\ +C,$$

or $$2I=x\sqrt{(x^2+a^2)}+a^2\ ln\ |\ \sqrt{(x^2+a^2)}+x\ |\ +C,$$

or $$I=\tfrac{1}{2}\ x\sqrt{(x^2+a^2)}+\tfrac{1}{2}\ a^2\ ln\ |\ \sqrt{(x^2+a^2)}+x\ |\ +C',$$

where we have written C' for $\frac{1}{2}$ C.

Aliter. Put $x=a \tan\theta$. Then $dx=a \sec^2\theta\ d\theta$.

$$I=\int \sqrt{(a^2\tan^2\theta+a^2)}\ .\ a\sec^2\theta\ d\theta,$$

$$=a^2\int \sec\theta\ .\ \sec^2\theta\ d\theta. \qquad ...(1)$$

Integrating by parts, taking first function $=\sec\theta$ and second function $=\sec^2\theta$, we have

$$I=a^2\left[\sec\theta\ .\ \tan\theta-\int (\sec\theta\ \tan\theta)\ .\ \tan\theta\ d\theta\right],$$

$$=a^2\left[\sec\theta\tan\theta-\int \sec\theta\ (\sec^2\theta-1)\ d\theta\right],$$

$$=a^2\left[\sec\theta\tan\theta-\int \sec^3\theta\ d\theta+\int \sec\theta\ d\theta\right],$$

$$=a^2\sec\theta\tan\theta-I+a^2\ ln\ |\ \sec\theta+\tan\theta\ |\ +C,$$

or $$2I=a^2\sec\theta\tan\theta+a^2\ ln\ |\ \sec\theta+\tan\theta\ |\ +C,$$

or $$I=x\sqrt{(x^2+a^2)}+\tfrac{1}{2}\ a^2\ ln\ |\ \sqrt{(x^2+a^2)}+x\ |\ +C',$$

after substituting back $\sec\theta=x/a$, and writing C' for the constant $-\frac{1}{2}\ a^2\ ln\ |\ a\ |+\frac{1}{2}C$.

Example 31. *Integrate* :

(*a*) $\int \sqrt{(16-x^2)}\,dx$; (*b*) $\int \sqrt{(x^2-16)}\,dx$;

(*c*) $\int \sqrt{(x^2+16)}\,dx$.

Solution.

(*a*) The integrand is of the form $\sqrt{(a^2-x^2)}$, where $a=4$.

$$\therefore \int \sqrt{(16-x^2)}\,dx=\tfrac{1}{2}\ x\ \sqrt{(16-x^2)}+8\sin^{-1}(x/4)+C.$$

(*b*) The integrand is of the form $\sqrt{(x^2-a^2)}$, where $a=4$.

$$\therefore \int \sqrt{(x^2-16)}\,dx=\tfrac{1}{2}\ x\ \sqrt{(x^2-16)}-8\ .\ ln\ |\ \sqrt{(x^2-16)}+x\ | +C.$$

(*c*) The integrand is of the from $\sqrt{(x^2+a^2)}$, where $a=4$.

$$\therefore \int \sqrt{(x^2+16)}\, dx = \tfrac{1}{2} x \sqrt{(x^2+16)} + 8 \ln \left| \sqrt{(x^2+16)} + x \right| + C.$$

5 4 3. Integrals of the type

$$\int \sqrt{(ax^2+bx+c)}\, dx, \int (px+q) \sqrt{(ax^2+bx+c)}\, dx$$

(A) Let $I = \int \sqrt{(ax^2+bx+c)}\, dx$

We have already seen as to how $\int \frac{dx}{\sqrt{(ax^2+bx+c)}}$ can be reduced to one of the standard forms

$$\int \frac{dx}{\sqrt{(a^2-x^2)}}, \int \frac{dx}{\sqrt{(x^2 \pm a^2)}}.$$

If we proceed exactly in the same manner, we can reduce I to one of the standard forms

$$\int \sqrt{(a^2-x^2)}\, dx, \int \sqrt{(x^2 \pm a^2)}\, dx.$$

(B) With the same notation as in the discussion of

$$\int \frac{px+q}{\sqrt{(ax^2+bx+c)}}\, dx,$$

we can get $I = \int (px+q) \sqrt{(ax^2+bx+c)}\, dx$

$$= l \int (2ax+b) \sqrt{(ax^2+bx+c)}\, dx + mJ$$

$$= \frac{2l}{3} (ax^2+bx+c)^{\frac{3}{2}} + mJ,$$

where $J = \int \sqrt{(ax^2+bx+c)}\, dx$ can be integrated as in (A) above.

Recall that $l = p/(2a)$, $m = q - bp/(2a)$.

The following examples will illustrate the method.

Example 32. *Evaluate* :

(*a*) $\int (4x+1) \sqrt{(x^2-x-2)}\, dx$

(*b*) $\int (3x+2) \sqrt{(1+4x-x^2)}\, dx.$

Solution.

(*a*) The derivative of x^2-x-2 (the expression under the radical sign)$=2x-1$. Therefore we write $4x+1=2(2x-1)+3$.

Then $$I=\int\left[2(2x-1)+3\right]\sqrt{(x^2-x-2)}\,dx,$$

$$=2\int(2x-1)\sqrt{(x^2-x-2)}\,dx+3\int\sqrt{(x^2-x-2)}\,dx,$$

$$=\frac{4}{3}(x^2-x-2)^{\frac{3}{2}}+3\int\left[\left(x-\frac{1}{2}\right)^2-\left(\frac{3}{2}\right)^2\right]^{\frac{1}{2}}dx,$$

$$=\frac{4}{3}(x^2-x-2)^{\frac{3}{2}}$$

$$+3\,ln\left|\left\{\left(x-\frac{1}{2}\right)^2-\left(\frac{3}{2}\right)^2\right\}^{\frac{1}{2}}+\left(x-\frac{1}{x}\right)\right|+C,$$

$$=\frac{4}{3}(x^2-x-2)^{\frac{3}{2}}+3\,ln\left|(x^2-x-2)^{\frac{1}{2}}+x-\frac{1}{2}\right|+C.$$

(*b*) Put $t=x-2$ (why ?). Then $x=t+2$; $dx=dt$,

$$I=\int(3x+2)\sqrt{(1+4x-x^2)}\,dx,$$

$$=\int\{3(t+2)+2\}\sqrt{\{1+4(t+2)-(t+2)^2\}}\,dt,$$

$$=\int(3t+8)\sqrt{(5-t^2)}\,dt,$$

$$=3\int t\sqrt{(5-t^2)}\,dt.$$

$$=3\int t\sqrt{(5-t^2)}\,dt+8\int\sqrt{(5-t^2)}\,dt,$$

$$=-3(5-t^2)^{\frac{3}{2}}+4\left[t\sqrt{(5-t^2)}+5\sin^{-1}t/\sqrt{5}\right]+C,$$

$$=-3(1+4x-x^2)^{\frac{3}{2}}+4(x-2)\sqrt{(1+4x-x^2)}$$
$$+20\sin^{-1}(x-2)/\sqrt{5}+C.$$

Remark. Try to compare the two methods used above.

EXERCISE 5 (*g*)

Evaluate :

1. $\int\sqrt{(9-x^2)}\,dx.$
2. $\int\sqrt{(x^2+7)}\,dx.$
3. $\int\sqrt{(4x^2+9)}\,dx.$
4. $\int\sqrt{(16-25x^2)}\,dx.$

5. $\int \sqrt{(x^2-25)}\, dx.$ 6. $\int \sqrt{(16x^2-49)}\, dx.$

7. $\int \sqrt{(4x^2-7)}\, dx.$ 8. $\int x \sqrt{(x^4+1)}\, dx.$

9. $\int \sqrt{(1+2x-3x^2)}\, dx.$ 10. $\int \sqrt{(3x^2-4x+1)}\, dx.$

11. $\int \sqrt{(2x-5)} \sqrt{(2+3x-x^2)}\, dx.$

12. $\int x \sqrt{(1+x-x^2)}\, dx.$

5·5. PARTIAL FRACTIONS AND THEIR USE IN INTEGRATION

We are already familiar with the process of combining a given number of fractions into a single fraction. For example,

$$\frac{2}{x+3}+\frac{1}{x+7}=\frac{3x+17}{(x+3)(x+7)} \qquad ...(1)$$

The L.H.S. when simplified gives the R.H.S. The reverse process, by which a given fraction can be resolved (or decomposed) into simpler fractions, also yields a unique result. The simpler fractions obtained by decomposing a given fraction are known as the partial fractions of the given fraction. In the present section we shall study the method of resolving a given fraction into partial fractions and then use it to integrate rational functions.

Resolution into Partial Fractions. Let the given fraction be $\frac{X}{Y}$, where X and Y are polynomials with real coefficients, in a variable X. We shall always assume that $\frac{X}{Y}$ is a proper fraction, for if that be not the case, we may always have

$$\frac{X}{Y}=Q+\frac{R}{Y}, \qquad ...(2)$$

where Q is a polynomial and $\frac{R}{Y}$ is a proper fraction.

Having put $\frac{X}{Y}$ in the form (2), we break up Y into simplest factors, *i.e.*, irredubible polynomials with real coefficients. (In the remainder of this section, the word 'factors' will always stand for irreducible polynomials with real coefficients unless stated otherwise). Three different cases now arise, according as the denominator Y of the given fraction consists of ;

(*i*) non-repeated linear factors only ;

(*ii*) linear factors only, some of which are repeated ;

(*iii*) linear as well as quadratic factors (which cannot be further resolved into real linear fractors) some of which may be repeated also.

It can be shown that

(*i*) to any non-repeated linear factor $(x+a)$ in the denominator of the given proper fraction, there corresponds a partial fraction $\frac{A}{x+a}$;

(*ii*) to any linear factor $(x+a)$ repeated twice in the denominator, there correspond two partial fractions $\frac{A}{x+a}$ and $\frac{B}{(x+a)^2}$; in general, the number of partial fractions corresponding to a factor $(x+a)$ repeated r times, we have r partial fractions

$$\frac{A}{x+a}, \frac{B}{(x+a)^2}, \ldots, \frac{K}{(x+a)^r}.$$

(*iii*) to any quadratic factor (ax^2+bx+c), which cannot be further resolved into linear factors, there corresponds a partial fraction $\frac{Ax+B}{ax^2+bx+c}$; to any twice repeated quadratic factor ax^2+bx+c (which cannot be further resolved into linear factors) in the denominator, there correspond two partial fractions $\frac{Ax+B}{ax^2+bx+c}$ and $\frac{Cx+D}{(ax^2+bx+c)^2}$, and so on.

The basic problem in resolving a given fraction into partial fractions is to find the values of the constants A, B, C, D,.......etc. (some of which may be equal to zero also). The technique of evaluating the constants in various cases will be illustrated by the following examples.

5·5·1. Case I : When the denominator consists of non-repeated linear factors only.

Example 33. *Resolve* $\frac{x+1}{(x-2)(x-3)}$ *into partial fractions and hence integrate* $\frac{x+1}{(x-2)(x-3)}$.

Solution. Let $\frac{x+1}{(x-2)(x-3)} \equiv \frac{A}{x-2} + \frac{B}{x-3}$ ...(*i*)

Multiplying both sides of (i) throughout by $(x-2)(x-3)$, we have

$$x+1 \equiv A(x-3)+B(x-2) \quad \ldots(ii)$$

Putting $x=2$ in both sides of the identity (ii), we have

$$3=-A, \text{ so that } A=-3. \quad \ldots(iii)$$

Again, putting $x=3$ in both sides of the identity (ii), we have

$$4=B, \text{ so that } B=4. \quad \ldots(iv)$$

Substituting the values of A and B in (i), we have

$$\frac{x+1}{(x-2)(x-3)}=\frac{4}{x-3}-\frac{3}{x-2}. \quad \ldots(v)$$

Integrating both sides of (v), we have

$$\int \frac{x+1}{(x-2)(x-3)}\,dx=\int \frac{4}{x-3}\,dx-\int \frac{3}{x-2}\,dx,$$

$$=4\,ln\,|\,x-3\,|-3\,ln\,|\,x-2\,|+C.$$

Note : The sign $\equiv$ connecting two rational functions indicates that they are identically equal *i.e.*, they have equal values for all values of the variable.

Example 34. *Integrate* $\dfrac{x^4+x^3-6x^2-13x-6}{x^3-7x-6}$ *by first resolving it into partial fractions.*

Since the degree of the numerator is not less than the degree of the denominator, we divide the numerator by the denominator. We then have

$$\frac{x^4+x^3-6x^2-13x-6}{x^3-7x-6} \equiv x+1+\frac{x^2}{x^3-7x-6}. \quad \ldots(1)$$

Now $\dfrac{x^2}{x^3-7x-6}$ is a proper fraction which can be resolved into partial fractions provided we can factorize the denominator.

Since $x^3-7x-6 \equiv (x+1)(x+2)(x-3)$, therefore

$$\frac{x^2}{x^3-7x-6}=\frac{x^2}{(x+1)(x+2)(x-3)} \quad \ldots(ii)$$

Let $$\frac{x^2}{(x+1)(x+2)(x-3)} \equiv \frac{A}{x+1}+\frac{B}{x+2}+\frac{C}{x-3}. \quad \ldots(iii)$$

Multiplying both sides of (iii) throughout by $(x+1)(x+2)(x-3)$, we have

$$x^2 \equiv A(x+2)(x-3)+B(x+1)(x-3)+C(x+1)(x+2) \quad \ldots(iv)$$

Putting $x=-1,\ -2,\ 3$ in succession in (iv), we have

$$A=-\frac{1}{4},\ B=\frac{4}{5},\ C=\frac{9}{20}.$$

Hence the given fraction

$$\equiv x+1-\frac{1}{4(x+1)}+\frac{4}{5(x+2)}+\frac{9}{20(x-3)}. \quad \ldots(v)$$

Integrating (v), we have

$$\int \frac{x^4+x^3-6x^2-13x-6}{x^3-7x-6}\,dx$$

$$=\int\left[x+1-\frac{1}{4(x+1)}+\frac{4}{5(x+1)}+\frac{9}{20(x-3)}\right]dx,$$

$$=\frac{1}{2}\,x^2+x-\frac{1}{4}\,\ln\,(x+1)+\frac{4}{5}\,\ln\,(x+2),$$

$$+\frac{9}{20}\,\ln\,(x-3)+C.$$

Example 35. *Resolve into partial fractions and integrate :*

$$\int \frac{x^4}{(x^2+1)(x^2+9)(x^2+16)}dx$$

Solution. Here the factors in the denominator are quadratic factors which cannot be further resolved into real linear factors. But we can simplify the problem by observing that only even powers of x occur in the given fraction. Putting $x^2=y$ in the given fraction, we have

$$\frac{x^4}{(x^2+1)\,(x^2+9)(x^2+16)}=\frac{y^2}{(y+1)\,(y+9)\,(y+16)} \quad \ldots(i)$$

Let $$\frac{y^2}{(y+1)\,(y+9)\,(y+16)}=\frac{A}{y+1}+\frac{B}{y+9}+\frac{C}{y+16} \quad \ldots(ii)$$

Multiplying both sides by $(y+1)(y+9)(y+16)$, we have

$$y^2\equiv A(y+9)(y+16)+B(y+1)(y+16)+C(y+1)(y+9) \quad \ldots(iii)$$

Putting $y=-1,\,-9,\,-16$ successively in (iii), we have

$$1=120A, \text{ or } \quad A=\frac{1}{120};$$

$$81=-56B, \text{ or } \quad B=-\frac{81}{56};$$

$$256=105C, \text{ or } \quad C=\frac{256}{105}.$$

Hence the given fraction

$$\equiv\frac{1}{120(y+1)}-\frac{81}{56(y+9)}+\frac{256}{105(y+16)},$$

$$\equiv\frac{1}{120(x^2+1)}-\frac{81}{56(x^2+9)}+\frac{256}{105(x^2+16)},$$

$\therefore$ The desired integral

$$=\int\left[\frac{1}{120(x^2+1)}-\frac{81}{56(x^2+9)}+\frac{256}{105(x^2+16)}\right]dx,$$

$$=\frac{1}{120}\tan^{-1}x-\frac{27}{56}\tan^{-1}\frac{x}{3}+\frac{64}{105}\tan^{-1}\frac{x}{4}+C$$

Example 36. *Integrate*

$$\frac{(x-1)(x-2)(x-3)}{(x-4)(x-5)(x-6)}$$

by first resolving it into partial fractions.

Solution. As the degree of the numerator is not less than the degree of the denominator, we have to divide the numerator by the denominator in order to reduce the given fraction to a proper fraction. Without performing the actual process of division it is obvious that the quotient will be 1 (Why ? Because the highest degree term in the numerator is x^3 and the highest degree term in the denominator is also x^3).

$$\therefore\quad \frac{(x-1)(x-2)(x-3)}{(x-4)(x-5)(x-6)}\equiv 1+\frac{A}{x-4}+\frac{B}{x-5}+\frac{C}{x-6}, \quad ...(i)$$

where the constants A, B and C have to be found.

Multiplying throughout by $(x-4)(x-5)(x-6)$, we have

$$(x-1)(x-2)(x-3)\equiv(x-4)(x-5)(x-6)+A(x-5)(x-6)$$
$$+B(x-4)(x-6)+C(x-4)(x-5). \quad ...(ii)$$

Putting $x=4$ in (*ii*), we get $3.2.1=2A$, or $A=3$.

Putting $x=5$ in (*ii*), we get $4.3.2=(-1)B$ or $B=-24$.

Putting $x=6$ in (*ii*), we get $5.4.3=2C$, or $C=30$.

Hence the given fraction

$$\equiv 1+\frac{3}{x-4}-\frac{24}{x-5}+\frac{30}{x-6}.$$

The desired integral

$$=\int\left(1+\frac{3}{x-4}-\frac{24}{x-5}+\frac{30}{x-6}\right)dx,$$

$$=x+3\ln|x-4|-24\ln|x-5|+30\ln|x-6|+C.$$

Note. Whenever the numerator and the denominator are of the same degree, the quotient can always be written out by inspection. The above method is then very useful.

EXERCISE 5 (*h*)

Resolve into partial fractions and integrate :

1. $\dfrac{2x+7}{(x+5)(x+3)}$.
2. $\dfrac{x+1}{(2-x)(x-3)}$.
3. $\dfrac{12x^2-30x+19}{(2x-1)(2x-3)(2x-5)}$.
4. $\dfrac{9x^2+7x-20}{x^3+2x^2-5x-6}$.
5. $\dfrac{3x^2-5x+1}{x^2-2x-3}$.
6. $\dfrac{4x^3+8x^2+4x-11}{(x-1)(2x+3)}$.
7. $\dfrac{x^4-10x^2+20x+13}{x^3+3x^2-x-3}$.
8. $\dfrac{x^3-5}{x^2-3x+2}$.
9. $\dfrac{(x-2)(x-3)(x-4)}{(2x-1)(x+1)(x-5)}$.
10. $\dfrac{x^2+1}{(x^2+2)(x^2+3)}$.
11. $\dfrac{4x^4+3}{(x^2+1)(x^2+3)(x^2+4)}$.
12. $\dfrac{x^2}{(x-a)(x-b)}$.
13. $\dfrac{x^2}{(x-a)(x-b)(x-c)}$.
14. $\dfrac{px^2+qx+r}{(x-a)(x-b)(x-c)}$.
15. $\dfrac{(x-a)(x-b)}{(x-c)(x-d)}$.
16. $\dfrac{x^2}{(x^2-a^2)(x^2-b^2)}$.

5·5·2. Case II : When the denominator consists of linear factors only, one or more of which may be repeated.

Example 37. *Resolve into partial fractions and integrate*

$$\frac{2x+1}{(x+2)(x-3)^2}.$$

Solution. Here the linear factor $(x-3)$ is repeated twice and $(x+2)$ is a non-repeated linear factor.

$$\text{Let}\quad \frac{2x+1}{(x+2)(x-3)^2}\equiv\frac{A}{x+2}+\frac{B}{x-3}+\frac{C}{(x-3)^2} \qquad \ldots(i)$$

Multiplying (*i*) throughout by $(x+2)(x-3)^2$, we have

$$2x+1\equiv A(x-3)^2+B(x+2)(x-3)+C(x+2) \qquad \ldots(ii)$$

Putting $x=-2$ in (*ii*), we have

$$-3=25A,\ \text{or}\ A=-\frac{3}{25}. \qquad \ldots(iii)$$

Putting $x=3$ in (*ii*), we have $7=5C$ or $C=\dfrac{7}{5}$. $\qquad \ldots(iv)$

In order to find B, we equate the coefficients of like powers of x on both sides of (*ii*). Since we have to find only one constant, let us consider only one of the relations that can be obtained in this

manner. Therefore, equating the constant terms on both sides of (*ii*), we have

$$1=9A-6B+2C. \quad \ldots(v)$$

Substituting the values of A and C in (*v*), we have $B=\frac{3}{25}$.

Hence the given fraction

$$\equiv -\frac{3}{25}\frac{1}{x+2}+\frac{3}{25}\frac{1}{x-3}+\frac{7}{5(x-3)^2}.$$

The desired integral

$$=\int\left[-\frac{3}{25}\frac{1}{x+2}+\frac{3}{25}\frac{1}{x-3}+\frac{7}{5(x-3)^2}\right]dx,$$

$$=-\frac{3}{25}\ln|x+2|+\frac{3}{25}\ln|x-3|-\frac{7}{5(x-3)}+c.$$

Note. In the above example the values of A and C have been found by giving special values to x in (*ii*), and the value of B has been obtained by equating rhe constant term on both sides of (*ii*), We could have found the values of all the theree constants A, B and C by equating the coefficients of like powers of x on both sides of (*ii*). But in order to do so, we would have to write three equations to determine A, B and C. This method will be explained in Example 40.

Example 38. *Evaluate*

$$\int\frac{4x-7}{(x-2)^3(x-3)}\,dx.$$

Solution. We shall first resolve the integrand into partial fractions. As the linear factor $(x-2)$ is repeated thrice, let

$$\frac{4x-7}{(x-2)^3(x-3)}\equiv\frac{A}{x-3}+\frac{B}{x-2}+\frac{C}{(x-2)^2}+\frac{D}{(x-2)^3} \quad \ldots(i)$$

Multiplying both sides of (*i*) by $(x-2)^3(x-3)$, we have

$$4x-7\equiv A(x-2)^3+B(x-3)(x-2)^2+C(x-3)(x-2)+D(x-3) \quad \ldots(ii)$$

Putting $x=3$ in (*ii*), we have $5=A$.

Putting $x=2$ in (*ii*), we have $1=-D$.

$\therefore \quad A=5,\ D=-1. \quad \ldots(iii)$

In order to obtain B and C, we equate coefficients of like powers of x on both sides of the identity (*ii*). Since the coefficients of the highest powers of x and the constant term in the identity can be written by inspection and they contain B as well as C, we have

$$\left.\begin{aligned}0&=A+B\\-7&=-8A-12B+6C-3D\end{aligned}\right\} \quad \ldots(iv)$$

From (*iv*) and (*v*) we have

$$B=-5,\ C=-5$$

Hence the given fraction

$$\equiv\frac{5}{x-3}-\frac{5}{x-2}-\frac{5}{(x-2)^2}-\frac{1}{(x-2)^3}.$$

The desired integral

$$=\int\left[\frac{5}{x-3}-\frac{5}{x-2}-\frac{5}{(x-2)^2}-\frac{1}{(x-2)^3}\right]dx,$$

$$=5\ln|x-3|-5\ln|x-2|$$

$$+\frac{5}{x-2}+\frac{1}{2(x-2)^2}+c.$$

Example 39. *Evaluate*

$$\int\frac{x^2}{(x-1)^3(x+1)}\,dx.$$

Solution. We shall first resolve the integrand into partial fractions. Let us put the repeated factor $x-1=y$, so that $x=1+y$. Substituting $x=1+y$ in the given fraction we find that the given fraction can be re-written as

$$\frac{(1+y)^2}{y^3(2+y)}=\frac{1}{y^3}\cdot\frac{1+2y+y^2}{2+y},$$

where we have arranged the terms in ascending powers of y.

Now we divide the numerator $1+2y+y^2$ by $2+y$ and carry on the division till the lowest power of y in the remainder is the same as that of the repeated factor y^3.

$$\begin{array}{r|l} & \frac{1}{2}+\frac{3}{4}y+\frac{1}{8}y^2 \\ 2+y & 1+2y+y^2 \\ & 1+\frac{1}{2}y \\ \hline & \frac{3}{2}y+y^2 \\ & \frac{3}{2}y+\frac{3}{4}y^2 \\ \hline & \frac{1}{4}y^2 \\ & \frac{1}{4}y^2+\frac{1}{8}y^3 \\ \hline & -\frac{1}{8}y^3 \end{array}$$

Thus the given fraction

$$=\frac{1}{y^3}\cdot\frac{1+2y+y^2}{2+y}=\frac{1}{y^3}\left[\frac{1}{2}+\frac{3}{4}y+\frac{1}{8}y^2-\frac{\frac{1}{8}y^3}{2+y}\right],$$

$$=\frac{1}{2y^3}+\frac{3}{4y^2}+\frac{1}{8y}-\frac{1}{8(2+y)},$$

$$=\frac{1}{2(x-1)^3}+\frac{3}{4(x-1)^2}+\frac{1}{8(x-1)}-\frac{1}{8(x+1)},$$

where we have replaced y by $x-1$ after obtaining the partial fractions.

The desired integral

$$=\int\left[\frac{1}{2(x-1)^3}+\frac{3}{4(x-1)^2}+\frac{1}{8(x-1)}-\frac{1}{8(x+1)}\right]dx,$$

$$=-\frac{1}{4(x-1)^2}-\frac{3}{4(x-1)}+\frac{1}{8}\ln|x-1|$$

$$-\frac{1}{8}\ln|x+1|+C.$$

Note. The above example can also be solved by the method of Example 38 (and example 38 could also have been solved by this method). The method explained above is, in fact, always applicable whenever there is one non-repeated linear factor and one repeated linear factor in the denominator.

Example 40. *Resolve* $\frac{x}{(x^2-1)^2}$ *into partial fractions and use them to evaluate*

$$\int\frac{x}{(x^2-1)^2}\,dx.$$

Solution. The denominator of the given fraction can be fractorized as $(x-1)^2(x+1)^2$. Here we have two linear factors, both of which are repeated. We, therefore, write

$$\frac{x}{(x^2-1)^2}=\frac{x}{(x-1)^2(x+1)^2},$$

$$\equiv\frac{A}{x-1}+\frac{B}{(x-1)^2}+\frac{C}{x+1}+\frac{D}{(x+1)^2}.$$

Multiplying both sides by $(x-1)^2(x+1)^2$, we have

$$x\equiv A(x-1)(x+1)^2+B(x+1)^2+C(x+1)(x-1)^2+D(x-1)^2 \qquad ...(i)$$

Equating the coefficients of x^3 on both sides of (i), we have

$$0=A+C \qquad ..(ii)$$

Equating the coefficients of x^2 on both sides of (i), we have

$$0=A+B-C+D \qquad ...(iii)$$

Again, equating the co-efficients of x on both sides of (i), we have

$$1=-A+2B-C-2D \qquad ...(iv)$$

Finally, equating the constant terms on both sides of (i), we have

$$0=-A+B+C+D \qquad ...(v)$$

We solve the four equations (ii)—(v) to obtain the values of A, B, C and D.

Subtracting both sides of (*v*) from the corresponding sides of (*iii*), we have

$$0=2A-2C, \text{ i.e., } A=C. \quad ...(vi)$$

From (*ii*) and (*vi*) we have $A=C=0$. ...(*vii*)

Substituting the values of A and C in (*iv*) and (*v*), we have

$$\left.\begin{aligned} 2B-2D&=1, \\ B+D&=0 \end{aligned}\right\} \quad ...(viii)$$

From (*viii*), we have

$$B=\tfrac{1}{4},\ D=-\tfrac{1}{4}.$$

Now $A=0$, $B=\frac{1}{4}$, $C=0$, $D=-\frac{1}{4}$. Therefore the given fraction

$$=-\frac{1}{4(x-1)^2}-\frac{1}{4(x+1)^2}.$$

The desired integral

$$=\int\left[\frac{1}{4(x-1)^2}-\frac{1}{4(x+1)^2}\right]dx$$

$$=-\frac{1}{4(x-1)}+\frac{1}{4(x+1)}+c.$$

Remarks 1. We could have obtained the integral of the function considered in the above example very easily by the substitution $x^2=t$. We would then get

$$I=\int\frac{x}{(x^2-1)^2}\,dx \qquad \left|\begin{array}{l}\text{Put } x^2=t \\ \therefore\ 2\,xdx=dt\end{array}\right.$$

$$=\frac{1}{2}\int\frac{dt}{(t-1)^2},$$

$$=-\frac{1}{2(t-1)}+c,$$

$$=-\frac{1}{2(x^2-1)}+c.$$

The technique of resolving a given rational function into partial fractions is important in its own right, because of its applications not only for finding integrals, but also for finding derivatives and elsewhere too. Suppose we were required to find the *derivative* of the given function.

It would have been cumbersome to find the derivative by the quotient rule.

2. The values of B and D could also have been obtained by putting $x=1$ and -1 successively in (*i*).

EXERCISE 5 (*i*)

Integrating the following :

1. $\dfrac{x}{(x+1)^2(x+1)}$.
2. $\dfrac{x^2+1}{(x-2)^2(x+3)}$.
3. $\dfrac{x^2+x+1}{(x-1)^2(x+1)}$.
4. $\dfrac{2x^2+x+1}{(x-1)^2(x+3)}$.
5. $\dfrac{x^2+2x+7}{(x-2)^2(x+3)}$.
6. $\dfrac{4x^3}{(x+1)^2(x^2-1)}$.
7. $\dfrac{x^2}{(x-1)^3(x+2)}$.
8. $\dfrac{x^3+x^2+x}{(x^2-1)(x+1)^3}$.
9. $\dfrac{x+4}{(x-2)^3(x+1)}$.
10. $\dfrac{x}{(x-1)^3(x-2)}$.
11. $\dfrac{3x+5}{(x-1)^3(x+2)}$.
12. $\dfrac{3x+5}{(x-1)^3(x+2)}$.
13. $\dfrac{1}{(x^2+x)(x^2-1)}$.
14. $\dfrac{x^3}{(x-2)^3(x+1)^2}$.
15. $\dfrac{x^2+a^2}{(x^2-a^2)(x+a)}$.

5·5·3. Case III : When the denominator consists of linear factors (some of which may be repeated also) as well as non-repeated quadratic factors (which cannot be further resolved into real linear factors).

Example 41. *Resolve into partial fractions and integrate :*

$$\frac{x}{(x^2+x-2)(x^2-x+2)}.$$

Solution. Observe that x^2+x-2 can be factorized as $(x+2)(x-1)$; however, x^2-x+2 cannot be factorized into real linear factors. We can therefore write the given fraction as

$$\frac{x}{(x+2)(x-1)(x^2-x+2)}.$$

Let $$\frac{x}{(x+2)(x-1)(x^2-x+2)}\equiv\frac{A}{x+2}+\frac{B}{x-1}+\frac{Cx+D}{x^2-x+2} \quad ...(i)$$

Multiplying both sides of (*i*) by $(x+2)(x-1)(x^2-x+2)$, we have

$$x\equiv A(x-1)(x^2-x+2)+B(x+2)(x^2-x+2)$$
$$+(Cx+D)(x+2)(x-1). \quad ...(ii)$$

Putting $x=-2$ in (*ii*), we have

$$-2=-24A,\ A=\frac{1}{12}. \quad ...(iii)$$

Putting $x=1$ in (*ii*), we have

$$1=6B, \text{ i.e., } B=\frac{1}{6}. \quad ...(iv)$$

In order to find C and D, we equate the coefficients of x^3 and the constant term on both sides of (*ii*). We, then, have

$$0=A+B+C. \quad ...(v)$$

$$0=-2A+4B-2D.$$

From (*iii*), (*iv*) and (*v*) we have

$$C=-\frac{1}{4}, D=\frac{1}{4}. \quad ...(vi)$$

Hence the given fraction

$$=\frac{1}{12(x+2)}+\frac{1}{6(x-1)}-\frac{x-1}{4(x^2-x+2)}$$

The desired integral

$$=\int\frac{1}{12(x+2)}\,dx+\int\frac{1}{6(x-1)}dx-\int\frac{(x-1)\,dx}{4(x^2-x+2)}.$$

$$=\frac{1}{12}\ln|x+2|+\frac{1}{6}\ln|x-1|-J, \quad ...(1)$$

where $$J=\frac{1}{4}\int\frac{(x-1)}{x^2-x+2}\,dx,$$

$$=\frac{1}{8}\int\frac{(2x-1)-1}{x^2-x+2}\,dx,$$

$$=\frac{1}{8}\ln|x^2-x+2|-\frac{1}{8}\int\frac{dx}{x^2-x+2},$$

$$=\frac{1}{8}\ln|x^2-x+2|-\frac{1}{8}\int\frac{dx}{\left(x-\frac{1}{2}\right)^2+\left(\frac{\sqrt{7}}{2}\right)^2},$$

$$=\frac{1}{8}\ln|x^2-x+2|-\frac{1}{8}\cdot\frac{2}{\sqrt{7}}\tan^{-1}\{(x-\tfrac{1}{2})/\sqrt{(7/2)}\}+c. \quad ...(2)$$

From (1) and (2) we find the desired integral

$$=\frac{1}{12}\ln|x+2|+\frac{1}{6}\ln|x-1|-\frac{1}{8}\ln|x^2-x+2|+\frac{1}{4\sqrt{7}}\tan^{-1}\{(2x-1)/\sqrt{7}\}+k.$$

where we have put k for $-c$.

Example 42. *Evaluate :*

$$\int\frac{x+a}{x^2(x-a)(x^2+a^2)}\,dx.$$

Solution. Let

$$\frac{x+a}{x^2(x-a)(x^2+a^2)} \equiv \frac{A}{x}+\frac{B}{x^2}+\frac{C}{x-a}+\frac{Dx+E}{x^2+a^2}$$

Multiplying both sides by $x^2(x-a)(x^2+a^2)$, we have

$$x+a \equiv Ax(x-a)(x^2+a^2)+B(x-a)(x^2+a^2)+Cx^2(x^2+a^2) +(Dx+E)x^2(x-a). \quad ...(i)$$

Putting $x=0$ in (i), we have $a=-a^3B$.

Again, putting $x=a$ in (i), we have $2a=Ca^2(2a^2)$.

$$\therefore \qquad B=-\frac{1}{a^2},\ C=\frac{1}{a^3} \qquad ...(ii)$$

Equating the coefficients x^4, x^3, x^2 on both sides of (i), we have

$$\left.\begin{aligned} 0&=A+C+D,\\ 0&=-aA+B+E-Da,\\ 0&=Aa^2-aB+a^2C-Ea. \end{aligned}\right\} \qquad ...(iii)$$

Substituting the values B and C from (ii) in (iii), we have

$$\left.\begin{aligned} A+D&=-\frac{1}{a^3},\\ aA-E+aD&=-\frac{1}{a^2},\\ a^2A-aE&=-\frac{2}{a}. \end{aligned}\right\} \qquad ...(iv)$$

From (iv), we have

$$D=\frac{1}{a^3},\ A=-\frac{2}{a^3},\ E=0. \qquad ...(v)$$

From (ii) and (v) we find that the given fraction

$$\equiv -\frac{2}{a^3x}-\frac{1}{a^2x^2}+\frac{1}{a^2(x-a)}+\frac{x}{a^3(x^2+a^2)}.$$

The desired integral

$$=-\frac{2}{a^3}\ln|x|+\frac{1}{a^2x}+\frac{1}{a^2}\ln|x-a| + \frac{1}{2a^3}\ln|x^2+a^2|+c.$$

Example 43. *Resolve into partial fractions and hence integrate*

$$\frac{x^3+x^2-x+1}{(x^2+1)(x-1)^2}.$$

Solution. Let

$$\frac{x^3+x^2-x+1}{(x^2+1)(x-1)^2} \equiv \frac{Ax+B}{x^2+1}+\frac{C}{x-1}+\frac{D}{(x-1)^2} \qquad ...(i)$$

Multiplying both sides by $(x^2+1)(x-1)^2$, we have

$$x^3+x^2-x+1\equiv(Ax+B)(x-1)^2+C(x-1)(x^2+1)+D(x^2+1). \quad ...(ii)$$

Putting $x=1$ in (*ii*), we have

$$2=2D, \text{ so that } D=1. \quad ...(iii)$$

Equating the coefficients of x^3, x^2 and the constant term on both sides of (*ii*), we have

$$\left.\begin{aligned} 1&=A+C,\\ 1&=-2A+B-C+D,\\ 1&=B-C+D. \end{aligned}\right\} \quad ...(iv)$$

Substituting the value of D in (*iv*), we have

$$A=0,\ C=1,\ B=1. \quad ...(v)$$

Hence the given fraction

$$=\frac{1}{x^2+1}+\frac{1}{x-1}+\frac{1}{(x-1)^2}.$$

$$\therefore \int \frac{x^3+x^2-x+1}{(x^2+1)(x-1)^2}\,dx=\tan^{-1}x+\ln|x-1|-\frac{1}{x-1}+c.$$

EXERCISE 5 (*j*)

Resolve into partial fractions and integrate :

1. $\dfrac{x}{x^3-1}$.
2. $\dfrac{x-1}{(x^2+4)(x-2)}$.
3. $\dfrac{3x-2}{(x^2+2)(x+2)}$.
4. $\dfrac{9x+7}{(x+3)(x^2+1)}$.
5. $\dfrac{x^2+1}{x^3+1}$
6. $\dfrac{43x+13}{(4x^2+3)(5-3x)}$.
7. $\dfrac{x^5}{x^4-1}$.
8. $\dfrac{x^2-2}{(x^2+1)(x^2-9)}$.
9. $\dfrac{x}{(x^3-1)(x+2)}$.
10. $\dfrac{x-4}{(x^2-3x+2)(x^2+4)}$.
11. $\dfrac{x^2+2x+2}{(x^2-1)(x^2+4)}$.
12. $\dfrac{1}{(x+1)^2(x^2+1)}$.
13. $\dfrac{x^2}{(x^2+1)(x-1)}$.

5·5·4. Integration of $\int \frac{dx}{a+b\cos x}$.

Let $$I=\int \frac{dx}{a+b\cos x}.$$

Three different cases arise :

Case I. $b=a$.

$$\mathrm{I}=\int \frac{dx}{a+a\cos x}$$

$$=\frac{1}{2a}\int \sec^2\left(\frac{x}{2}\right)dx$$

$$=\frac{1}{a}\tan(x/2)+\mathrm{C}.$$

Case II. $b=-a$.

$$\mathrm{I}=\int \frac{dx}{a-a\cos x}$$

$$=\frac{1}{2a}\int \csc^2 \frac{x}{2}\, dx$$

$$=-\frac{1}{a}\cot(x/2)+c.$$

Case III. $b^2 \neq a^2$.

Put $t=\tan(x/2)$. Then $dt=\frac{1}{2}\sec^2(x/2)\,dx$,

so that $$dx=2\frac{dt}{1+\tan^2(x/2)}=2\frac{dt}{1+t^2}.$$

$$\therefore\quad \mathrm{I}=\int \frac{dx}{a+b\cos x}=\int \frac{1}{a+\dfrac{b(1-t^2)}{1+t^2}}\cdot\frac{2\,dt}{1+t^2}$$

$$=2\int \frac{dt}{a(1+t^2)+b(1-t^2)}$$

$$=2\int \frac{dt}{(a-b)t^2+a+b}$$

$$=-\frac{2}{a-b}\int \frac{dt}{t^2+\dfrac{a+b}{a-b}}.$$

Two different possibilities arise :

(*i*) $a^2>b^2$. In this case,

$$\frac{a+b}{a-b}=\frac{(a+b)^2}{a^2-b^2}>0,$$

$$\therefore\quad \mathrm{I}=\frac{2}{a-b}\left[\sqrt{\left(\frac{a-b}{a+b}\right)}\tan^{-1}\left(t\sqrt{\left(\frac{a-b}{a+b}\right)}\right)\right]+\mathrm{C},$$

$$=-\frac{2}{\sqrt{(a^2-b^2)}}\tan^{-1}\left(\sqrt{\left(\frac{a-b}{a+b}\right)}\tan\frac{x}{2}\right)+\mathrm{C}.$$

(*ii*) $a^2 < b^2$. In this case,

$$\frac{a+b}{a-b}=\frac{(a+b)^2}{a^2-b^2}<0$$

$$\therefore \quad I=\frac{2}{a-b}\cdot\frac{1}{2}\sqrt{\left(\frac{b+a}{b-a}\right)}\ln\left(\frac{t-\sqrt{\left(\frac{b+a}{b-a}\right)}}{t+\sqrt{\left(\frac{b+a}{b-a}\right)}}\right)+C,$$

$$=\frac{1}{\sqrt{(b^2-a^2)}}\ln\left(\frac{\tan\frac{x}{2}+\sqrt{\left(\frac{b+a}{b-a}\right)}}{\tan\frac{x}{2}-\sqrt{\left(\frac{b+a}{b-a}\right)}}\right)+C.$$

5·5·5. Integrals of the type $\int\frac{dx}{a+b\sin x}$.

Three different cases arise :

Case I. $b=a$.

In this case,

$$I=\int\frac{dx}{a+a\sin x},$$

$$=\frac{1}{a}\int\frac{dx}{1-\cos\left(\frac{\pi}{2}+x\right)},$$

$$=\frac{1}{2a}\int\csc^2\left(\frac{\pi}{4}+\frac{x}{2}\right)dx,$$

$$=-\frac{1}{2a}\cot\left(\frac{\pi}{4}+\frac{x}{2}\right)+C.$$

Case II. $b=-a$.

In this case,

$$I=\int\frac{dx}{a-a\sin x},$$

$$=\frac{1}{a}\int\frac{dx}{1+\cos\left(\frac{\pi}{2}+x\right)},$$

$$=\frac{1}{2a}\int\sec^2\left(\frac{\pi}{4}+\frac{x}{2}\right)dx,$$

$$=\frac{1}{2a}\tan\left(\frac{\pi}{4}+\frac{x}{2}\right)+C.$$

Case III. $b^2 \neq a^2$.

In this case we put $\tan(x/2)=t$,

so that

$$\frac{1}{2}\sec^2\left(\frac{x}{2}\right)dx=dt,$$

or

$$\frac{1}{2}(1+t^2)dx=dt,$$

or

$$dx=\frac{2\,dt}{1+t^2}.$$

Also, $\sin x=\dfrac{2\tan(x/2)}{1+\tan^2(x/2)}=\dfrac{2t}{1+t^2}.$

$$\therefore \quad I=\int\frac{dx}{a+b\sin x}=\int\frac{1}{a+b\cdot\dfrac{2t}{1+t^2}}\cdot\frac{2\,dt}{1+t^2},$$

$$=2\int\frac{dt}{at^2+2bt+a},$$

$$=\frac{2}{a}\int\frac{dt}{t^2+\dfrac{2bt}{a}+1},$$

$$=\frac{2}{a}\int\frac{dt}{\left(t+\dfrac{b}{a}\right)^2+\dfrac{a^2-b^2}{a^2}}.$$

There are two different possibilities :

(*i*) $a^2-b^2>0$. In this case

$$I=\frac{2}{a}\,\frac{a}{\sqrt{(a^2-b^2)}}\tan^{-1}\left\{\left(t+\frac{b}{a}\right)\frac{a}{\sqrt{(a^2-b^2)}}\right\}+C,$$

$$=\frac{2}{\sqrt{(a^2-b^2)}}\tan^{-1}\frac{a\tan(x/2)+b}{\sqrt{(a^2-b^2)}}+c.$$

(*ii*) $a^2-b^2<0$. In this case,

$$I=\frac{2}{a}\cdot\frac{a}{2\sqrt{(b^2-a^2)}}\ln\left|\frac{t+\dfrac{b}{a}-\dfrac{\sqrt{(b^2-a^2)}}{a}}{t+\dfrac{b}{a}+\dfrac{\sqrt{(b^2-a^2)}}{b}}\right|+C,$$

$$=\frac{1}{\sqrt{(b^2-a^2)}}\ln\left|\frac{a\tan\dfrac{x}{2}+b-\sqrt{(b^2-a^2)}}{a\tan\dfrac{x}{2}+b+\sqrt{(b^2-a^2)}}\right|+C.$$

Remarks 1. By using the substitution

$x=\pi/2+y$, we find that

$$\int \frac{dx}{a+b\sin x}=\int \frac{dy}{a+b\cos y},$$

so that the above integral reduces to an earlier one.

2. In actual practice we do not use the results obtained above but proceed afresh by using the substitution $\tan(x/2)=t$.

3. The substitution $\tan(x/2)=t$ can be used whenever the integrand is of the form $F(\cos x, \sin x)$.

Example 44. *Evaluate* $\int \frac{dx}{1-\cos\alpha\cos x}$.

Solution. Let us put $\tan x/2=t$, so that $x=2\tan^{-1} t$,

$$dx=\frac{2}{1+t^2}=dt.$$

Also, $\cos x=\frac{1-t^2}{1+t^2}$.

Then

$$I=\int \frac{dx}{1-\cos\alpha\cos x}$$

$$=\int \frac{1}{1-\frac{\cos\alpha\,(1-t^2)}{1+t^2}}\cdot\frac{2\,dt}{1+t^2}$$

$$=\int \frac{2\,dt}{(1+t^2)-\cos\alpha\,(1-t^2)}$$

$$=\int \frac{2dt}{(1+\cos\alpha)t^2+(1-\cos\alpha)},$$

$$=\frac{1}{\cos^2\frac{\alpha}{2}}\int \frac{dt}{t^2+\tan^2(\alpha/2)},$$

$$=\frac{1}{\cos^2(\alpha/2)}\cdot\frac{1}{\tan(\alpha/2)}\tan^{-1}\left(\frac{t}{\tan(\alpha/2)}\right)+C$$

$$=\frac{2}{\sin\alpha}\tan^{-1}(\tan x/2)\cot(\alpha/2)+C.$$

Example 45. *Evaluate :*

(*a*) $\int \frac{dx}{5+4\cos x}$ (*b*) $\int \frac{dx}{4+5\cos x}$

Solution.

(*a*) $I=\int \frac{dx}{5+4\cos x}$

$$=\int \frac{dx}{5(\cos^2 x/2+\sin^2 x/2)+4(\cos^2 x/2-\sin^2 x/2)},$$

$$=\int \frac{dx}{9\cos^2 x/2+\sin^2 x/2},$$

$$=\int \frac{\sec^2 x/2\, dx}{\tan^2 x/2+9}. \qquad ...(1)$$

Putting $\tan x/2=t$, $\frac{1}{2}\sec^2 x/2\, dx=dt$ in (1), we have

$$I=2\int \frac{dt}{t^2+9}=\frac{2}{3}\tan^{-1}(t/3)+C,$$

$$=\frac{2}{3}\tan^{-1}(\tfrac{1}{3}\tan(x/2))+C.$$

(*b*) $$I=\int \frac{dx}{4+5\cos x},$$

$$=\int \frac{dx}{4(\cos^2 x/2+\sin^2 x/2)+5(\cos^2 x/2-\sin^2 x/2)},$$

$$=\int \frac{dx}{9\cos^2 (x/2)-\sin^2 (x/2)},$$

$$=\int \frac{\sec^2 (x/2)\, dx}{9-\tan^2 (x/2)}. \qquad ...(1)$$

Putting $\tan x/2=t$, $\frac{1}{2}\sec^2 x/2\, dx=dt$ in (1), we have

$$I=2\int \frac{dt}{9-t^2},$$

$$=\frac{1}{3}\ln\left|\frac{3+t}{3-t}\right|+C,$$

$$=\frac{1}{3}\ln\left|\frac{3+\tan x/2}{3-\tan x/2}\right|+C.$$

Remark. You must have noted that the technique used above is slightly different from the one described in section 5·5·4. We could as well have proceeded in exactly the same manner as outlined there.

Example 46. *Evaluate* :

(*a*) $\int \frac{dx}{5+4\sin x}$ (*b*) $\int \frac{dx}{4+5\sin x}$.

Solution.

(*a*) $$I=\int \frac{dx}{5(\cos^2 x/2+\sin^2 x/2)+4\sin x/2\cos x/2)},$$

$$=\int \frac{\sec^2 x/2\, dx}{5+5\tan^2 x/2+8\tan x/2}. \qquad ...(1)$$

Putting $\tan x/2=t$, $\frac{1}{2}\sec^2 x/2\, dx=dt$ in (1), we have

$$I=\int\frac{2\,dt}{5+8t+5t^2},$$

$$=\frac{2}{5}\int\frac{dt}{t^2+\frac{8}{5}t+1},$$

$$=\frac{2}{5}\int\frac{dt}{\left(t+\frac{4}{5}\right)^2+\left(\frac{3}{5}\right)^2},$$

$$=\frac{2}{5}\left(\frac{5}{3}\tan^{-1}\frac{t+4/5}{3/5}\right)+C.$$

$$=\tfrac{2}{3}\tan^{-1}\frac{5t+4}{3}+C,$$

$$=\tfrac{2}{3}\tan^{-1}\tfrac{1}{3}(5\tan(x/2)+4)+C.$$

(*b*) $$I=\int\frac{dx}{4(\cos^2(x/2)+\sin^2(x/2))+5.2\sin(x/2)\cos(x/2)},$$

$$=\int\frac{\sec^2(x/2)dx}{4+10\tan(x/2)+4\tan^2(x/2)} \qquad ...(1)$$

Putting $\tan(x/2)=t$, $\frac{1}{2}\sec^2(x/2)dx=dt$ in (1), we have

$$I=\int\frac{2dt}{4+10t+4t^2},$$

$$=\tfrac{1}{2}\int\frac{dt}{t^2+\frac{5}{2}t+1},$$

$$=\tfrac{1}{2}\int\frac{dt}{(t+\frac{5}{4})^2-(\frac{3}{4})^2},$$

$$=\tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{4}{3}\ln\left|\frac{(t+\frac{5}{4})-\frac{3}{4}}{(t+\frac{5}{4})+\frac{3}{4}}\right|+C,$$

$$=\tfrac{1}{3}\ln\left|\frac{2t+1}{2t+4}\right|+C,$$

$$=\tfrac{1}{3}\ln\left|\frac{2\tan x/2+1}{2\tan x/2+4}\right|+C.$$

Remark. Compare the above working with that described in section 5·5·4.

EXERCISE 5 (*k*)

Integrate the following :

1. $\dfrac{1}{3+2\cos x}$, 2. $\dfrac{1}{2+\cos x}$.

3. $\dfrac{1}{5+3\cos x}$. 4. $\dfrac{1}{1+2\cos x}$.

5. $\dfrac{1}{4+3\cos x}$. 6. $\dfrac{1}{13+12\cos x}$.

7. $\dfrac{1}{5+12\cos x}$. 8. $\dfrac{1}{1-2\sin x}$.

5·6. Integration of $\sin^m x \cos^n x$

The integral of $\sin^m x \cos^n x$ occurs in many applications of integral calculus. In this section we shall see as to how we can integrate this function when m and n are non-negative integers.

(*i*) If m is an odd positive integer, say $2k+1$, then we put $\cos x=t$. Then $-\sin x\, dx=dt$,

$$I=\int \sin^m x \cos^n x\, dx,$$

$$=\int \sin^{2k+1} x \cos^n x\, dx,$$

$$=-\int (1-t^2)^k t^n\, dt,$$

which can be easily evaluated by first expanding $(1-t^2)^k$ by the binomial theorem.

(*ii*) If n is an odd positive integer, then we put $\sin x=t$ and proceed as above.

(*iii*) If both m and n are odd, we may proceed either as in (*i*) or as in (*ii*).

(*iv*) If both m and n are even, then we generally use the method of successive reduction which we describe below :

By using the technique of integration by parts we obtain a relation between

(*a*) $\int \sin^m x \cos^n x\, dx$ and $\int \sin^{m-2} x \cos^n x\, dx$ and also (if necesary) a relation between

(*b*) $\int \sin^m x \cos^n x\, dx$ and $\int \sin^m x \cos^{n-2} x\, dx$.

Such relations are called reduction formulae.

By repeated application of reduction formulae the integral

$$I_{m,n}=\int \sin^m x \cos^n x\, dx$$

can be expressed in terms of $I_{0,0}$ which can be integrated immediately. The method of successive reduction can be used in cases (*i*)—(*iii*) also.

The following examples will illustrate the method :

Example 47. *Evaluate* :

(*a*) $\int sin^5 x\, dx$ (*b*) $\int \cos^7 x\, dx.$

Solution.

(*a*) Let $I=\int \sin^5 x\, dx.$

Put $\cos x=t.$ Then $-\sin x\, dx=dt$

$$I=\int(1-\cos^2 x)^2 \sin x\, dx,$$

$$=-\int(1-t^2)^2\, dt,$$

$$=-\int(1-2t^2+t^4)dt,$$

$$=-\frac{t^5}{5}+\frac{2t^3}{3}-t+C,$$

$$=-\frac{\cos^5 x}{5}+\frac{2\cos^3 x}{3}-\cos x+C.$$

(*b*) $I=\int \cos^7 x\, dx$

Put $\sin x=t.$ Then $\cos x\, dx=dt.$

$$I=\int(1-\sin^2 x)^3 \cos x\, dx,$$

$$=\int(1-t^2)^3\, dt,$$

$$=\int(1-3t^2+3t^4-t^6)dt,$$

$$=-\frac{t^7}{7}+\frac{3t^5}{5}-t^3+t+C,$$

$$=-\frac{\sin^7 x}{7}+\frac{3\sin^5 x}{5}-\sin^3 x+\sin x+C.$$

Example 48. *Evaluate* :

(*a*) $\int sin^3 x\, cos^4 x\, dx.$ (*b*) $\int sin^4 x\, cos^3 x\, dx.$

Solution.

(*a*) $I=\int \sin^3 x \cos^4 x\, dx.$

The exponent of sin x is odd. We shall put $\cos x=t$.

Then $\quad -\sin x\,dx=dt$.

$$I=-\int(1-t^2)t^4\,dt,$$

$$=-\frac{t^5}{5}+\frac{t^7}{7}+C,$$

$$=-\frac{\cos^5x}{5}+\frac{\cos^7x}{7}+C.$$

(*b*) $$I=\int\sin^4x\cos^3x\,dx.$$

The exponent of cos x is odd. We shall put $\sin x=t$. Then $\cos x\,dx=dt$.

$$I=\int t^4(1-t^2)dt,$$

$$=\frac{t^5}{5}-\frac{t^7}{7}+C,$$

$$=\frac{\sin^5x}{5}-\frac{\sin^7x}{7}+C.$$

Example 49. *For each positive integer m, let* $I_m=\int\sin^m x\,dx$. *Show that* $I_m=-\frac{\cos x\sin^{m-1}x}{m}+\frac{m-1}{m}I_{m-2}$. *Use it to evaluate* I_6.

Solution.

$$I_m=\int\sin^{m-1}.\sin x\,dx$$

Taking first function$=\sin^{m-1}x$, second function$=\sin x$, and integrating by parts, we have

$$I_m=-\sin^{m-1}x\cos x-\int(m-1)\sin^{m-2}x\cos x(-\cos x)dx$$

$$=-\sin^{m-1}x\cos x+(m-1)\int\sin^{m-2}x(1-\sin^2x)dx$$

$$=-\sin^{m-1}x\cos x+(m-1)[I_{m-2}-I_m],$$

or $$mI_m=-\sin^{m-1}x\cos x+(m-1)I_{m-2},$$

or $$I_m=-\frac{1}{m}\sin^{m-1}x\cos x+\frac{(m-1)}{m}I_{m-2} \qquad \text{...(1)}$$

Putting $m=6, 4, 2$ in (1), we have

$$I_6=-\tfrac{1}{6}\sin^5x\cos x+\tfrac{5}{6}I_4$$

$$I_4=-\tfrac{1}{4}\sin^3x\cos x+\tfrac{3}{4}I_2$$

$$I_2=-\tfrac{1}{2}\sin x\cos T+\tfrac{1}{2}I_0.$$

Also, $$I_0=\int \sin^0 x\,dx=\int dx=x+C.$$

From the above relations we can find I_2, I_4, I_6. After calculations we find that

$$I_6=-\tfrac{1}{6}\sin^5 x\cos x-\tfrac{5}{24}\sin^3 x\cos x-\tfrac{5}{16}\sin x\cos x$$
$$+\tfrac{5}{16}x+C', \text{ where } C' \text{ is an arbitrary constant.}$$

Remark. If we were simply to evaluate $\int \sin^6 x\,dx$ we could have first expressed $\sin^6 x$ as a sum of sines/cosines of multiples of x and then integrated each term of the sum to obtain the result. The The result would have been obtained in a different but equivalent form. In that case we would have proceeded as follows :

Since $$\sin^2 x=\tfrac{1}{2}(1-\cos 2x),$$
therefore $$\sin^6 x=\tfrac{1}{8}(1-\cos 2x)^3,$$
$$=\tfrac{1}{8}[1-3\cos 2x+3\cos^2 2x-\cos^3 2x] \quad ...(i)$$

Writing $$\cos^2 2x=\tfrac{1}{2}(1+\cos 4x),$$
$$\cos^3 2x=\tfrac{1}{4}(\cos 6x+3\cos 2x),$$

(i) becomes
$$\sin^6 x=\tfrac{1}{8}[1-3\cos 2x+\tfrac{3}{2}(1+\cos 4x)-\tfrac{1}{4}(\cos 6x+3\cos 2x)\}$$
$$=\frac{1}{32}[10-15\cos 2x+6\cos 4x-\cos 6x] \quad ...(ii)$$

From (ii) we have
$$\int \sin^6 x\,dx=\frac{1}{32}[10x-\frac{15}{2}\sin 2x$$
$$+\frac{3}{2}\sin 4x-\frac{1}{6}\sin 6x]+C \quad ...(iii)$$

If the exponents are not too big, this method also works well.

Example 50. *For each positive integer n, let $J_n=\int \cos^n x\,dx$. Show that*

$$J_n=\frac{1}{n}\cos^{n-1}x\sin x+\frac{n-1}{n}J_{n-2}.$$

Use it to evaluate J_6.

Solution.

$$J_n=\int \cos^{n-1}x\,.\cos x\,dx$$

Taking first function$=\cos^{n-1}x$, second function$=\cos x$, and integrating by parts, we have

$$J_n=\cos^{n-1}\sin x-\int(n-1)\cos^{n-2}x\,(-\sin x)\,.\,(\sin x)dx,$$

$$=\cos^{n-1}x\ \sin x+(n-1)\int\cos^{n-2}x\,(1-\cos^2x)dx,$$

$$=\cos^{n-1}x\sin\ x+(n-1)(J_{n-2}-J_n),$$

or $$nJ_n=\cos^{n-1}x\ \sin x+(n-1)J_{n-2},$$

$$J_n=\frac{1}{n}\cos^{n-1}x\ \sin x+\frac{n-1}{n}J_{n-2}. \qquad \ldots(1)$$

Putting $n=6, 4, 2$ in succession in (1) we have

$$J_6=\tfrac{1}{6}\cos^5x\ \sin x+\tfrac{5}{6}J_4,$$
$$J_4=\tfrac{1}{4}\cos^3 x\ \sin x+\tfrac{3}{4}J_2,$$
$$J_2=\tfrac{1}{2}\cos\ x\ \sin x+\tfrac{1}{2}J_0.$$

Also, $J_0=\int\cos^0x\,dx=\int dx=x+C_1.$

From the above relations we find that

$$J_2=\tfrac{1}{2}\cos x\ \sin x+\tfrac{1}{2}\ x+C_2,$$
$$J_4=\tfrac{1}{4}\cos^3 x\ \sin x+\tfrac{3}{4}(\tfrac{1}{2}\cos x\ \sin x+\tfrac{1}{2}\ x)+C_3,$$
$$=\tfrac{1}{4}\cos^3 x\ \sin x+\tfrac{3}{8}\cos x\ \sin x+\tfrac{3}{8}\ x+C_3,$$
$$J_6=\tfrac{1}{6}\cos^5 x\ \sin x+\tfrac{5}{24}\cos^3 x\ \sin x+\tfrac{5}{16}\cos x\ \sin x$$
$$+\tfrac{5}{16}\ x+C,$$

where C, C_1, C_2, C_3 are all arbitrary constants.

Remarks 1. We could have deduced the result of this example from that of the preceding example (or vice-versa) by the substitution, $x=y+\pi/2$. In fact by substituting $x=y+\pi/2$ and $dx=dy$ in the relation

$$\int\sin^m x\,dx=-\frac{\cos x\ \sin^{m-1}x}{m}+\frac{m-1}{m}\int\sin^{m-2}x dx,$$

and remembering that $\sin(y+\pi/2)=\cos y$,

$\cos(y+\pi/2)=-\sin y$, we have

$$\int\cos^m y\,dy=\frac{\sin y\cos^{m-1}y}{m}+\frac{m-1}{m}\int\cos^{m+2}y\,dy,$$

which is the reduction formula proved above (except that we have 'n' instead of 'm').

2. It would be interesting to express $\cos^6x$ as a sum of cosines of multiples x by using the formula $\cos^2 x=\frac{1}{2}(1+\cos 2x)$ and then integrating term by term.

Example 51. *For every pair of non-negative integers m and n, let* $I_{m,n}=\int \sin^m x \cos^n x \, dx$.

Show that

(*a*) $(m+n)\, I_{m,n}=-\sin^{m-1} x \cos^{n+1} x+(m-1)\, I_{m-2,n}$.

(*b*) $(m+n)\, I_{m,n}=\sin^{m+1}x \cos^{n-1}x+(n-1)\, I_{m,n-2}$.

Use the above reduction formulae to evaluate

$$I_{4,2}=\int \sin^4 x \cos^2 x \, dx.$$

Solution. (*a*) By taking first function$=\sin^{m-1} x$ and second function$=\sin x \cos^n x$, and integrating by parts, we have

$$I_{m,n}=\int(\sin^{m-1}x)(\sin x \cos^n x)\, dx,$$

$$=\sin^{m-1}x\left(-\frac{\cos^{n+1}x}{n+1}\right)-\int(m-1)\sin^{m-2}x \cos x \left(-\frac{\cos^{n+1}x}{n+1}\right) dx.$$

Multiplying throughout by $n+1$ and replacing $\cos^2 x$ by $1-\sin^2 x$ in the integrand on the right hand side, we have

$$(n+1)\, I_{m,n}=-\sin^{m-1}x \cos^{n+1}x+(m-1)\int \sin^{m-2}x\,(1-\sin^2 x) \cos^n x \, dx,$$

$$=-\sin^{m-1}x \cos^{n+1}x+(m-1)(I_{m-2,n}-I_{m,n}),$$

or $\quad (m+n)\, I_{m,n}=-\sin^{m-1}x \cos^{n+1}x+(m-1)\, I_{m-2,n} \quad$ ...(A)

(*b*) We can either proceed exactly as in (*a*) above or deduce the result from (A) as follows :

Interchanging *m* and *n* throughout in (A) above, we have

$$(n+m)\int \sin^n x \cos^m x \, dx=-\sin^{n-1}x \cos^{m+1}x+(n-1)\int \sin^{n-2}x \cos^m x \, dx \quad ...(1)$$

Put $x=y+\pi/2$, $dx=dy$ in (1) throughout, and use $\sin(y+\pi/2)=\cos y$, $\cos(y+\pi/2)=-\sin y$. Then

$$(m+n)\int \cos^n y[(-1)^m \sin^m y]\, dy$$

$$=-\cos^{n-1}y(-1)^{m+1} \sin^{m+1}y)$$

$$+(n-1)\int \cos^{n-2}y[(-1)^m \sin^m y] dy,$$

or $$(m+n)\int \sin^m y \cos^n y\, dy = \sin^{m+1} y \cos^{n-1} y$$

$$+(n-1)\int \sin^m y \cos^{n-2} y\, dx, \quad ...(B)$$

which is the desired formula (except that we have got the variable y everywhere instead of x, which we know to be immaterial).

To evaluate $I_{4,\,2}$ we use the reduction formula (A) twice taking $(m, n)=(4, 2)$ and $(2, 2)$ in succession. We then have

$$6\, I_{4,\,2} = -\sin^3 x \cos^3 x + 3I_{2,\,2}, \quad ...(2)$$

$$4\, I_{2,\,2} = -\sin x \cos^3 x + I_{0,\,2} \quad ...(3)$$

From (2) and (3) we have

$$6\, I_{4,\,2} = -\sin^3 x \cos^3 x - \frac{3}{4}\sin x \cos^3 x + \frac{3}{4} I_{0,\,2} \quad ...(4)$$

Using formula (B) for $m=0$, $n=2$, we have

$$2\, I_{0,\,2} = \sin x \cos x + I_{0,\,0} \quad ...(5)$$

Also $$I_{0,\,0} = x + C_1 \quad ...(6)$$

Using (5) and (6), we have from (4),

$$I_{4,\,2} = -\frac{1}{6}\sin^3 x \cos^3 x - \frac{1}{8}\sin x \cos^3 x$$

$$+\frac{1}{16}\sin x \cos x + \frac{1}{16}x + C.$$

Remarks. 1. Observe that we used formula (A) to reduce the exponent of $\sin x$ as much as we could, and then used (B) to reduce the exponent of $\cos x$. This can always be done till $I_{m,\,n}$ is expressed in terms of any one of the four integrals $I_{1,\,1}$, $I_{1,\,0}$, $I_{0,\,1}$, $I_{0,\,0}$. All these four integrals can be evaluated directly.

2. The formulae in Examples 49 and 50 are particular cases of the formulae (A) and (B) above.

In fact, putting $n=0$ in (A) throughout we have

$$m\, I_{m,\,0} = -\sin^{m-1} x \cos x + (m-1)\, I_{m-2,\,0}$$

which is the same as the formula of Example 49.

Similarly, by putting $m=0$ in (B) throughout we get the formula of Example 50.

EXERCISE 5 (*l*)

Evaluate :

1. $\int \cos^5 x\, dx$. 2. $\int \sin^7 x\, dx$.

3. $\int \sin^4 x\, dx.$ 4. $\int \cos^4 x\, dx.$

5. $\int \sin^2 x \cos^3 x\, dx.$ 6. $\int \sin^2 x \cos^4 x\, dx.$

7. $\int \sin^3 x \cos^5 x\, dx.$ 8. $\int \sin^6 x \cos^2 x\, dx.$

9. $\int \sin^4 x \cos^4 x\, dx.$ 10. $\int \sin^5 x \cos^3 x\, dx.$

TEST YOUR UNDERSTANDING V

In each of the following problems, four alternatives are given out of which only one is correct. Put a tick mark (✓) against the correct alternative :

1. A primitive of x^3 is
   (a) $3x^2$ (b) $\frac{1}{3}x^2$
   (c) $\frac{1}{4}x^4$ (d) x^4.
2. A primitive of e^{2x} is
   (a) $2xe^{2x}$ (b) $-2e^{2x}$
   (c) e^{x^2} (d) $\frac{1}{2}e^{2x}$.
3. $\int ln\, x\, dx$ equals
   (a) $ln\, ln\, x+C$ (b) $\frac{1}{x}+C$
   (c) $x\, ln\, x$ (d) $x\, ln\, x-x+C$
4. An indefinite integral of $\frac{1}{\sqrt{(4-x^2)}}$ is
   (a) $\sin^{-1}(2x)$ (b) $\frac{1}{2}\sin^{-1}(2x)$
   (c) $\sin^{-1}(x/2)$ (d) $2\sin^{-1}(2x)$.
5. An anti-derivative of $\tan x$ is
   (a) $ln \sec x$ (b) $ln \cos x$
   (c) $\sec^2 x$ (d) $\cot x$.
6. $\int \frac{dx}{a^2+x^2}$ equals
   (a) $ln\,(a^2+x^2)+C$ (b) $\frac{1}{a}\tan^{-1}(x/a)+C$
   (c) $\sin^{-1}(x/a)+C$ (d) $-2x/(a^2+x^2)^2$.
7. A primitive of $f(x)$ is $\sec^2 x$. Then f must be
   (a) $2\sec^2 x \tan x$ (b) $\tan x$
   (c) $\frac{1}{3}\sec^3 x$ (d) $2\sec x$.

8. A primitive of f is $ln\ \{\sqrt{(x^2-4)}+x\}$. f must be

(a) $\dfrac{1}{\sqrt{(x^2-4)}}$ (b) $\sqrt{(x^2-4)}$

(c) $\dfrac{1}{\sqrt{(x^2-4+x)}}$ (d) $\dfrac{2}{3}(x^2-4)^{3/2}$

9. A primitive of $(\sin x)/\cos^2 x$ equals

(a) $\tan x$ (b) $\sec x$

(c) $\cos^2 x/\sin x$ (d) $1/\cos^2 x$.

10. An anti-derivative of $f(x)$ is $\sin^{-1}(x^3)$. $f(x)$ is

(a) $3x^2 \sin^{-1}(x^3)$ (b) $\dfrac{3x^2}{\sqrt{(1-x^6)}}$

(c) $\cos^{-1} x^3$ (d) $3x^2(1-x^6)^{1/2}$.

REVIEW EXERCISE V

Evaluate the following integrals :

1. $\displaystyle\int \frac{x^3-1}{x^3+x}\,dx$. (*Roorkee Entrance, 1988*)

2. $\displaystyle\int \frac{1+x^{-2/3}}{1+x}\,dx$. (*Roorkee Entrance, 1987*)

3. $\displaystyle\int \frac{ln\ x}{x^3}\,dx$. (*Roorkee Entrance, 1986*)

4. $\displaystyle\int \frac{x^4}{(x-1)(x^2+1)}\,dx.$ (*Roorkee Entrance, 1986*)

5. $\displaystyle\int \frac{dx}{\sqrt{(2x+3)}-\sqrt{2x}}$. (*A.I.S.S.C.E., 198*)

6. $\displaystyle\int \frac{x^2+x-1}{x^2+x-6}\,dx.$ (*A.I.S.S.C.E., 1988*)

7. $\displaystyle\int \sqrt{\left(\frac{a-x}{a+x}\right)}\,dx.$ (*D.B.S.S.C.E., 1988*)

8. $\displaystyle\int e^x (a+be^x)^6\,dx.$ (*A.I.S.S.C.E., 1984*)

9. $\displaystyle\int \frac{-x^2+4x+3}{(x+2)(x-1)}\,dx.$ (*D.B.S.S.CE., 1987*)

10. $\displaystyle\int \frac{e^x(2-\sin 2x)}{1-\cos 2x}\,dx.$ (*D.B.S.S.C.E., 1989*)

11. $\displaystyle\int \frac{x \sin^{-1} x^2}{\sqrt{(1-x^4)}}\,dx.$ (*D.B.S.S.C.E., 1986*)

12. $\int \frac{1-\tan x}{1+\tan x}\, dx.$ (*D.B.S.S.C.E., 1984*)

13. $\int \frac{\cos x}{(\cos x/2+\sin x/2)^3}\, dx$ (*D.B.S.S.C.E., 1985*)

14. $\int \tan^{-1}\left(\frac{3x-x^3}{1-3x^2}\right) dx.$ (*D.B.S.S.C.E, 1984*)

15. $\int \frac{dx}{1+\sin x}.$ (*A.I.S.S.C.E., 1984*)

16. $\int \frac{dx}{\sqrt{(x+a)}+\sqrt{(x+b)}}$ (*A.S.S.C.E., 1984*)

17. $\int \frac{dx}{x \cos^2 (1+ln\, x)}$ (*A.I.S.S.C.E., 1989*)

18. $\int \left\{1+2 \tan x\, (\tan x+\sec x)\right\}^{1/2} dx.$ (*Roorkee Entrance, 1987*)

19. $\int (ln\, x)^2\, dx.$ (*A.I.S.S.C.E., 1984*)

20. $\int \frac{2x\, dx}{\sqrt{(1-x^2-x^4)}}$ (*A.I.S.S.C.E., 1984*)

21. $\int e^x (\tan x - ln \mid \sec x \mid)\, dx.$ (*D.B.S.S.C.E., 1988*)

22. $\int \frac{\sin 4x}{\sin x}\, dx.$ (*D.B.S.S.C.E., 1986*)

23. $\int \frac{2x\, dx}{(x^2+1)(x^2+2)}$ (*D.B.S.S.C.E., 1988*)

24. $\int \frac{\sin 2x\, dx}{a^2 \cos^2 x+b^2 \sin^2 x}$ (*D.B.S.S.C.E., 1989*)

25. $\int \frac{\sin \theta\, d\theta}{\sqrt{(1+\cos \theta)}}$ (*D.B.S.S.C.E., 1989*)

26. $\int \frac{1}{2} \sin 7x \cos 5x\, dx.$ (*A.I.S.S.C.E., 1988*)

27. $\int \sec x\, ln\, (\sec x+\tan x)\, dx.$ (*A.I.S.S.C.E., 1986*)

28. $\int e^x (\cot x-\csc^2 x)\, dx.$ (*A.I.S.S.C.E., 1986*)

9. $\int \frac{2x}{(2x+1)^2}\, dx.$

30. $\int \frac{dx}{4 \sin^2 x+5 \cos^2 x}$ (*A.I.S.S.C.E., 1986*)

SUMMARY

1. $\int [f(x) \pm g(x)]dx = \int f(x)dx \pm \int g(x)dx.$

2. $\int k f(x)dx = k \int f(x)dx.$

3. Let $f(x)$ be a function which has a primitive. If $x=\phi(t)$ is a derivable function of t, then

$$\int f(x)\, dx = \int f(\phi(t))\phi'(t)dt.$$

4. Let f and g be two fuuctions such that $\int g(x)dx$ and

$\int [f'(x) \int g(x)dx]\, dx$ exist. Then

$$\int f(x)g(x)dx = f(x)\int g(x)dx] - \int [f'(x)\int g(x)dx]dx.$$

5. The following standard integrals are often found useful :

(1) $\int x^n dx = \frac{x^{n+1}}{n+1} + C,\ n \neq -1.$

(2) $\int \frac{dx}{x} = \ln |x| + C,\ x \neq 0.$

(3) $\int \sin x\, dx = -\cos x + C.$

(4) $\int \cos x\, dx = \sin x + C.$

(5) $\int \tan x\, dx = \ln |\sec x| + C.$

(6) $\int \cot x\, dx = \ln |\sin x| + C.$

(7) $\int \sec x\, dx = \ln |\sec x + \tan x| + C.$

(8) $\int \csc x\, dx = \ln \left|\tan \frac{x}{2}\right| + C.$

(9) $\int \sec^2 x\, dx = \tan x + C.$

(10) $\int \csc^2 x\, dx = -\cot x + C.$

(11) $\int \sec x \tan x\, dx = \sec x + C.$

(12) $\int \csc x \cot x \, dx = -\csc x + C.$

(13) $\int e^x dx = e^x + C.$

(14) $\int a^x dx = \dfrac{a^x}{\ln a} + C, \; a > 0, \; a \neq 1.$

(15) $\int \dfrac{dx}{x^2 - a^2} = \dfrac{1}{2a} \ln \left| \dfrac{x-a}{x+2} \right| + C.$

(16) $\int \dfrac{dx}{a^2 - {}^2x} = \dfrac{1}{2a} \ln \left| \dfrac{a+x}{a-x} \right| + C.$

(17) $\int \dfrac{dx}{x^2 + a^2} = \dfrac{1}{a} \tan^{-1}(x/a) + C.$

(18) $\int \dfrac{dx}{\sqrt{(a^2 - x^2)}} = \sin^{-1}(x/a) + C.$

(19) $\int \dfrac{dx}{\sqrt{(x^2 + a^2)}} = \ln\{ \mid \sqrt{(x^2 + a^2)} + x \mid \} + C.$

(20) $\int \dfrac{dx}{\sqrt{(x^2 - a^2)}} = \ln\{ \mid \sqrt{(x^2 - a^2)} + x \mid \} + C.$

(21) $\int \sqrt{(a^2 - x^2)} dx = \dfrac{1}{2} x \sqrt{(a^2 - x^2)} + \dfrac{a^2}{2} \sin^{-1}(x/a) + C.$

(22) $\int \sqrt{(x^2 + a^2)} dx = \frac{1}{2} x \sqrt{(x^2 + a^2)} + \dfrac{a^2}{2} \ln\{ \mid \sqrt{(x^2 + a^2)} + x \mid \} + C.$

(23) $\int \sqrt{(x^2 - a^2)} dx = \frac{1}{2} x \; (x^2 - a^2) - \dfrac{a^2}{2} \ln\{ \mid \sqrt{(x^2 - a^2)} + x \mid \} + C.$

(24) $\int e^{ap} \cos bx \, dx = \dfrac{e^{ax}}{\sqrt{(a^2 + b^2)}} \cos (bx - \tan^{-1}(b/a)) + C.$

(25) $\int e^{ax} \sin bx \, dx = \dfrac{e^{ax}}{\sqrt{(a^2 + b^2)}} \sin (bx - \tan^{-1}(b/a)) + C.$

(26) If $\int f(t) dt = F(t)$, then

$$\int f(ax) + b) dx = \frac{1}{a} F(ax + b) + C.$$

(27) $\int [f(x)]^n f'(x)dx = \frac{1}{n+1}[f(x)]^{n+1} + C,\ n \neq -1.$

(28) $\int \frac{f'(x)}{f(x)}\,dx = \ln|f(x)| + C.$

(29) $\int e^x[f(x)+f'(x)]dx = e^x f(x) + C.$

(30) If $I_m = \int \sin^m x\,dx$, then

$$I_m = -\frac{\cos x \sin^{m-1} x}{m} + \frac{m-1}{m} I_{m-2}$$

(31) If $I_n = \int \cos^n x\,dx$, then

$$I_n = \frac{\cos^{n-1} x \sin x}{n} + \frac{n-1}{n} I_{n-2}.$$

(32) If $I_{m,\,n} = \int \sin^m x \cos^n x\,dx,$

then

$$(m+n)I_{m,n} = -\sin^{m-1}x\ \cos^{n+1} + (m-1)I_{m-2,n}.$$

$$= \sin^{m+1}x \cos^{n-1}x + (n-1)I_{m,\,n-2}.$$

HISTORICAL NOTE

The search for methods of integration may be traced back to the Greeks. The *method of exhaustion* used by *Archimedes* (287-212 B.C.) to find the area and arc-length of a circle, area of a parabolic segment, area of an ellipse, surface and volume of a sphere, was nothing else but integral calculus in a primitive form.

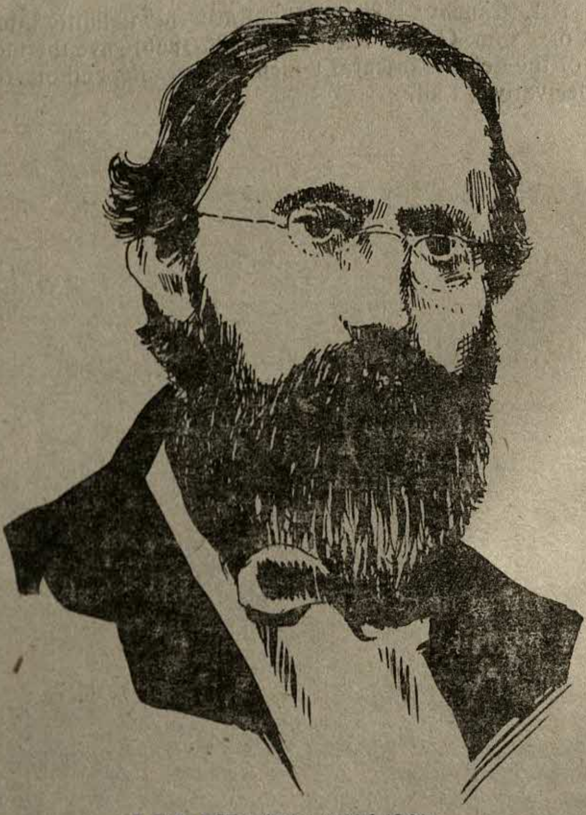
B. Cavalieri (1598-1647), a pupil of Galileo and Professor at Bologna (Italy), published his *method of indivisibles* in his book *Geometria indivisibilibus*, in 1635. Cavalieri solved most of the problems regarding determination of areas of planetary orbits posed by Kepler. Using this method with some modifications, important contributions were made by *P. Fermat* (1601-1655), *G. Roberval* (1602-1675), E. Torricelli (1608-1647), and B. Pascal (1623-1662).

John Wallis (1616-1703) in his *Arithmetica infinitorum* published in 1665, systematized the *method of indivisibles*. *I. Barrow* (1630-1677), Newton's teacher and predecessor at Cambridge made important contributions to integral calculus in his *Lectiones Geometricae* published

in 1670. The discovery of the calculus independently but about the same time by *Issac Newton* (1642-1727) and G. W. Leibnitz (1646-1717) was the consolidation and organization of effort made by many mathematicians over past several centuries.

A.L. Cauchy (1789-1857) defined the definite integral as the limit of a sum. G.F.B. Riemann (1826-1866) gave the modern definition of the definite integral which does not depend on the notion of the derivative at all.

□ □



G.F.B. RIEMANN (1826-1866)

Georg Friedrich Bernhard Riemann, one of the greatest geniuses of the nineteenth century, was a student of Karl Friedrich Gauss and Wilhelm Weber. In 1846 he went to Göttingen to study theology but soon changed over to mathematics. The concept of definite integral was put on a firm foundation by him. He made valuable contributions to several areas of mathematics. He died at the age of forty. His work has continued to inspire mathematicians even to this day.

CHAPTER 6

Definite Integrals

6·1. INTRODUCTION

In the previous chapter we defined integration as the inverse process of differentiation. However, historically the integral calculus was invented first and the differential calculus later. The concept of integration first arose in connection with determination of areas of plane regions bounded by curves and an integral was recognized as the limit of a certain sum. It was only later that Newton and Leibnitz established an intimate relationship between the processes of integration and differentiation, known now as the *fundamental theorem of integral calculus* which we shall discuss in this chapter. A definite integral will be defined as the limit of a sum and it will be shown how a definite integral can be used to define the area of some special region. The application of the fundamental theorem to determine the areas of several important geometric figures will then be considered.

6·1·1. The Area of a Plane Region

In elementary geometry, you must have come across formulae for finding the areas of many plane figures. For instance, the area of a rectangle is equal to the product of the lengths of two adjacent sides. Similarly, the area of a triangle is equal to half the product of the length of any side and the length of the altitude perpendicular to that side. One can determine areas of polygons and other plane figures bounded by straight lines in a similar manner. But what about areas of figures bounded by curved lines ? The area of such a figure is sometimes defined as the number of unit squares (squares with sides of length 1) which can fit in the figure. But this definition is very unsatisfactory. Consider, for example, a circle of radius 1. One would say that the irrational number π denotes the area of this circle. According to the above definition, π unit squares should fit into the circle. But what does 'π' squares mean ? Even if one considers a circle of radius $1/\sqrt{\pi}$, which should have area 1, it is difficult to believe that a unit square will fit in this circle, because it is impossible to divide the unit square into pieces which can be arranged to form a circle.

The ancient Greeks used a certain method which came to be known as the method of exhaustion to give meaning to the areas of

such regions. But it was Archimedes who made the most elegant applications of this method. For this reason perhaps, the method of exhaustion is sometimes erroneously attributed to Archimedes. Using this method, Archimedes proved that the area of a parabolic segment is 4/3 of the area of a triangle on the same base and having the same vertex. Let us now discuss the way this method is used for assigning a meaning to the area of a circle.

Consider a circle S and inscribe a regular polygon S_1 of n sides in it. Now, the area of the regular polygon S_1 is less than the area of the circle as some portions of the circle are left out. However, if we go on increasing the number n of sides of S_1, the area of S_1 becomes closer and closer to the area of the circle. It is intuitively clear that if n tends to infinity, the area of the polygon tends to become the area of the circle. Similarly let us circumscribe a regular polygon S_2 of n sides about the circle. The area of S_2 is always greater than the area of the circle, but as n, the number of sides, increases, the area of S_2 goes on becoming closer and closer to the area of S and as n tends to infinity, the area of S_2 tends to the area of the circle.

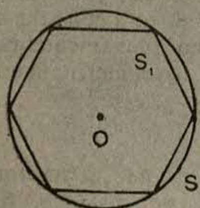


Fig. 6·1.

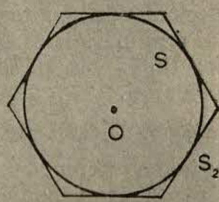


Fig. 6·2.

The method of exhaustion thus consists in approximating the given region by two simpler figures whose areas can be determined. The great astronomer Kepler, in his noteworthy work on planetary motions, asserted that a planet describes equal focal sectors of ellipses in equal times. This naturally demands some method of finding the areas of such sectors and the one invented by Kepler was called by him the method of 'the sum of the radii'—a rude kind of integration. Kepler considered solids as composed of infinitely many infinitely small cones or infinitely thin disks etc., the summation of which became later, the problem of integration. Kepler's attempt at integration led Cavaleiri to develop his method of indivisibles. He considered solids as composed of superposed surfaces, surfaces as made up of lines and lines as made up of points. He then proceeded to find lengths, areas and volumes by the summation of these indivisibles. Toricelli, Roberval, Fermat, Pascal and Wallis also obtained a number of results in this direction.

The method of exhaustion was transformed into a powerful branch of mathematics, now called integral calculus, largely through

the efforts of Newton and Leibnitz. The subject was put on a sound mathematical basis during the nineteenth century by Augustin Louis Cauchy (1789-1857) and George Bernhard Friedrich Riemann (1826-1866).

We shall now see for ourselves as to how the areas of certain regions can be viewed as the limits of certain sums. We shall make it precise below.

6·1·2. The area of a region as the limit of a sum

Suppose we are given a function f which is defined and continuous over the closed interval $[a, b]$. Let f be monotonic in $[a, b]$ and let $f(x) \geqslant 0$ for all $x \in [a, b]$.

Divide $[a, b]$ into n sub-intervals of equal length. If h be the length of each sub-interval, then $nh=b-a$. As n increases, h decreases. In fact as $n \to \infty$, $h \to 0$.

Let the n sub-intervals, be $[a, a+h]$, $[a+h, a+2h]$,............ $[a+\overline{n-1}\,h, a+nh]$. We shall consider two different cases according as $f(x)$ is increasing or decreasing.

Case I. Let $f(x)$ be increasing on $[a, b]$. On the sub-intervals $[a, a+h]$, $[a+h, a+2h]$,......$[a+\overline{n-1}\,h, a+nh]$ erect rectangles of heights $f(a), f(a+h)$,......, $f(a+\overline{n-1}\,h)$ respectively as shown in Fig. 6·3 (a). The sum S of the areas of these rectangles is given by

$$S=hf(a)+hf(a+h)+\dots\dots\dots+hf(a+\overline{n-1}\,h)$$
$$=h\{f(a)+f(a+h)+\dots\dots\dots+f(a+\overline{n-1}\,h)\} \qquad \dots(1)$$

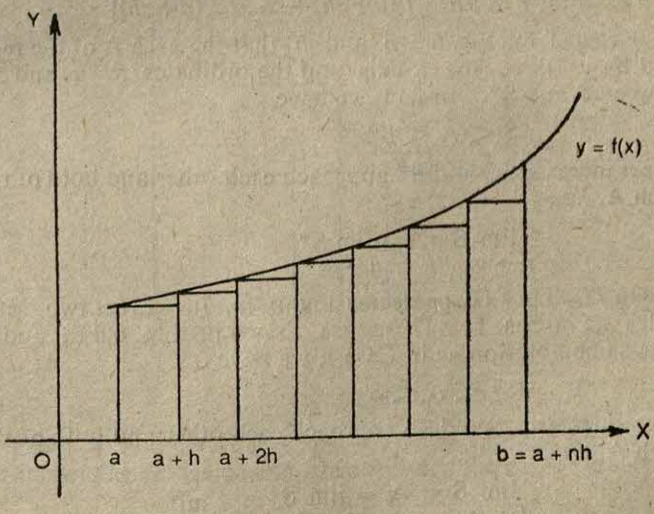


Fig. 6·3 (a).

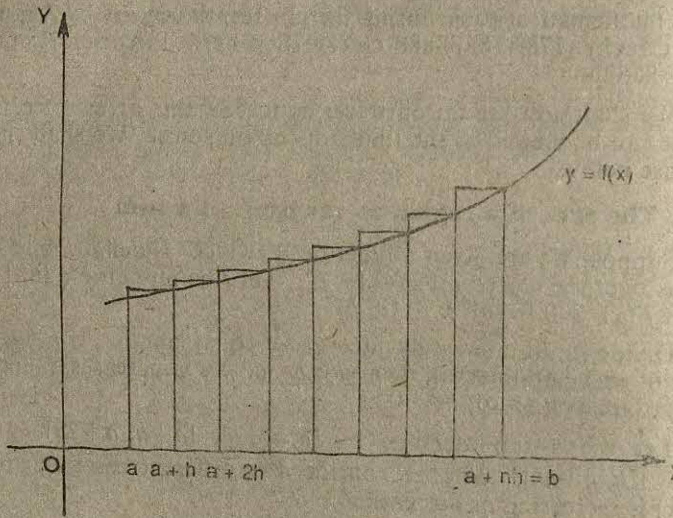


Fig. 6·3 (*b*).

Let us now erect rectangles of heights $f(a+h), f(a+2h), \ldots\ldots f(a+nh)$ on the sub-intervals $[a, a+h]$, $[a+h, a+2h], \ldots\ldots\ldots$ $[a+\overline{n-1}\,h, a+nh]$ respectively, as shown in Fig. 6·3 (*b*).

The sum S* of the areas of the rectangles is given by

$$S^* = hf(a+h) + hf(a+2h) + \ldots\ldots + hf(a+nh)$$

$$= h\{f(a+h) + f(a+2h) + \ldots\ldots + f(a+nh)\} \qquad \ldots(2)$$

It is clear from Fig. 6·3 (*a*) and (*b*) that the area A of the region bounded by $y=f(x)$, the x-axis and the ordinates $x=a$ and $x=b$ lies between S and S*. In fact, we have

$$S < A < S^* \qquad \ldots(3)$$

As n increases, S and S* approach each other and both of them approach A, *i.e.*,

$$\lim_{n \to \infty} S = A = \lim_{n \to \infty} S^* \qquad \ldots(4)$$

Case II. Let $f(x)$ be decreasing in $[a, b]$. Erect two sets of rectangles as in Case I. These are shown in Fig. 6·4 (*a*) and (*b*). With the same notation as in Case I, we have

$$S^* < A < S \qquad \ldots(4)$$

As n increases, S and S* approach each other and both of them approach A, *i.e.*,

$$\lim_{n \to \infty} S = A = \lim_{n \to \infty} S^* \qquad \ldots(5)$$

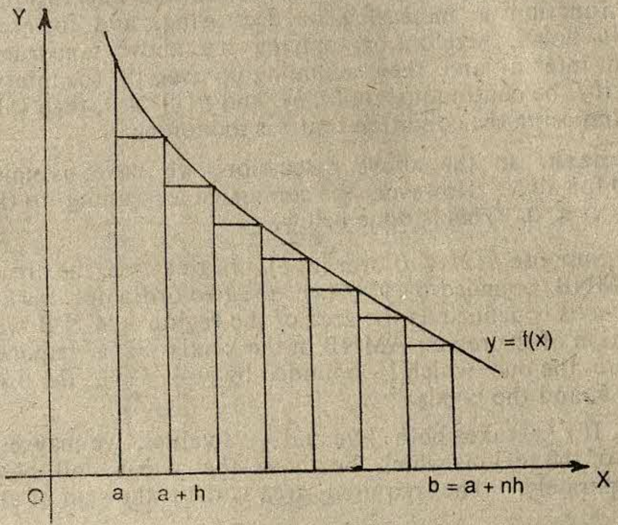


Fig. 6·4 (*a*).

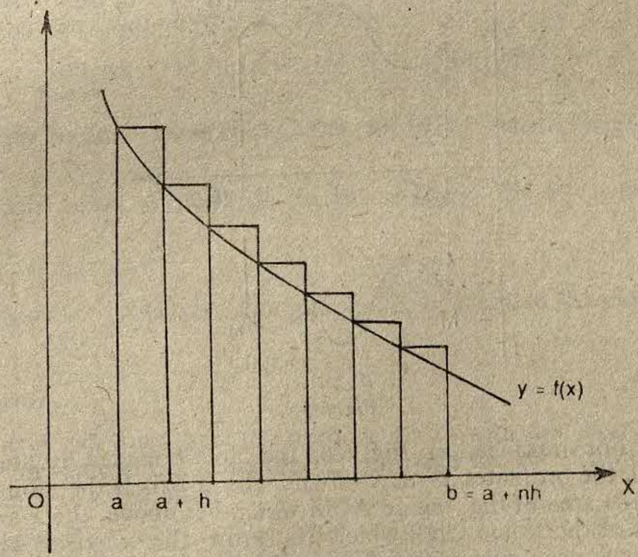


Fig. 6·4 (*b*).

From (4) and (5) we find that if f be monotonic on $[a, b]$, *i.e.*, either increasing or decreasing, then in both the cases we have

$$\lim_{n \to \infty} S = A = \lim_{n \to \infty} S^* \qquad \text{...(6)}$$

We have already seen as to how we can find intervals in which a given function is increasing or decreasing, and for each such interval (6) holds, therefore by applying the above construction to each such interval and then summing up over all the intervals we find that if f be continuous on $[a, b]$ and $f(x) \geqslant 0$, then (6) holds without imposing the condition that f is monotone.

Remark. In the above discussion, we have assumed that $f(x) \geqslant 0$ for all x. However, one can assign a meaning to the area even if $f(x) \leqslant 0$. This is done below.

(*i*) Suppose $f(x) \leqslant 0$ over $[a, b]$. In this case, the area of the region AMNB bounded by $y=f(x)$, the two ordinates $x=a$, $x=b$, and the x-axis is defined as the area of the region AM′N′B which is the reflection of the region AMNB in the x-axis. The required area is therefore the one which is bounded by $y=-f(x)$, the ordinates $x=a$, $x=b$, and the x-axis.

(*ii*) If $f(x)$ takes both +ve and −ve values, we may consider the areas of the regions which lie above the x-axis and below the x-axis separately. The required area is then the sum of all these

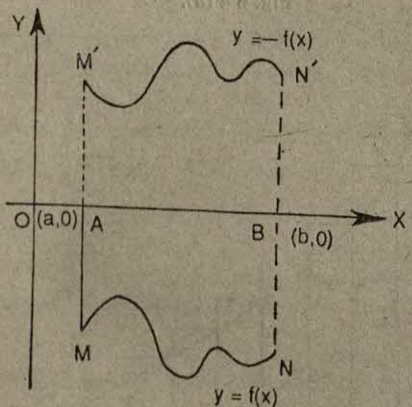


Fig. 6·5.

areas. For instance, in Fig. 6·6, the area bounded by the curve $y=f(x)$, the ordinates $x=a$, $x=b$ and the x-axis is the sum of two areas—the area of the region ACM which lies above the x-axis and the area of the region CNB which lies below the x-axis. The area CNB has a meaning as defined in (*i*) above.

6·2. THE DEFINITE INTEGRAL

We are now ready to give the definition of the definite integral. The definition, although motivated by the concept of area, does not depend upon it.

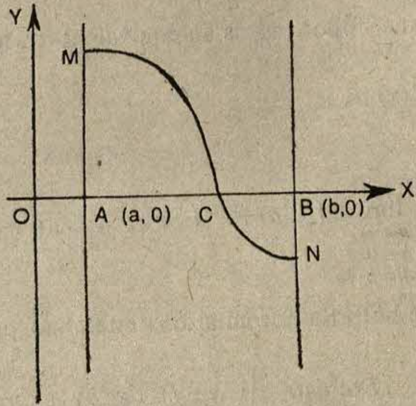


Fig. 6·6.

Definition 6·1. *Let $f(x)$ be defined and continuous over the closed interval $[a, b]$. Let n be a positive integer and h be a positive real number such that $nh=b-a$. Then*

$$\lim_{n \to \infty} h\,[f(a+h)+f(a+2h)+\dots+f(a+nh)]$$

is called the **definite integral** *of $f(x)$ over $[a, b]$ and is denoted by* $\int_a^b f(x)\,dx$. Here, a is called the **lower limit** and b is called the **upper limit** of $\int_a^b f(x)\,dx$. The interval $[a, b]$ is called the **range of integration.**

The world 'limit' here has nothing to do with the word limit as used in differential calculus. It only signifies the 'end points' of the range of integration.

Remarks 1. We shall often state the above definition in the form

$$\int_a^b f(x)\,dx = \lim_{\substack{n \to \infty \\ h \to 0 \\ nh=b-a}} h\,\{f(a+h)+f(a+2h)+\dots+f(a+nh)\}.$$

2. Since $\lim\limits_{h\to 0} hf(h)$ and $\lim\limits_{h\to 0} hf(a+nh)$ are both equal to zero, therefore the following is an equivalent formulation of the definition of $\int_a^b f(x)\,dx$:

$$\int_a^b f(x)\,dx = \lim_{\substack{n\to\infty \\ h\to 0 \\ nh=b-a}} h\,\{f(a)+f(a+h)+\cdots+f(a+\overline{n-1}\,h)\}.$$

We shall use both the formulations interchangeably.

Example 1. *Evaluate* $\int_a^b (2x+7)\,dx$ *by expressing it as the limit of a sum.*

Solution. Here $f(x)=2x+7$.

Therefore $\int_a^b (2x+7)\,dx$

$$=\lim_{\substack{h\to 0 \\ n\to\infty \\ nh=b-a}} h\{f(a+h)+f(a+2h)+\cdots+f(a+nh)\}$$

$$=\lim_{\substack{h\to 0 \\ n\to\infty \\ nh=b-a}} h\,[2(a+h)+7+2(a+2h)+7+\ldots+2(a+nh)+7]$$

$$=\lim_{\substack{h\to 0 \\ n\to\infty \\ nh=b-a}} h\,[(2a+7)\,n+(2h)\,(1+2+3+\ldots+n)]$$

$$=\lim_{\substack{h\to 0 \\ n\to\infty \\ nh=b-a}} [(2a+7)(nh)+h^2n\,(n+1)]$$

$$=\lim_{\substack{h\to 0 \\ h\to\infty \\ nh=b-a}} [(nh^2)+nh\,(2a+7)+(nh)\,.\,h]$$

$$=\lim_{h\to 0} [(b-a)^2+(b-a)\,(2a+7)+(b-a)\,h]$$

$$=(b-a)^2+(b-a)\,(2a+7)$$

$$=b^2-a^2+7\,(b-a).$$

Remark. Observe that we have tried to arrange the expression in terms of nh and then put $nh=b-a$ before taking limits.

Example 2. *Evaluate* $\int_a^b e^x\,dx$ *by expressing it as the limit of a sum.*

Solution. Here $f(x)=e^x$.

Therefore, $\int_a^b e^x\,dx$

$$=\lim_{\substack{h\to 0\\ n\to\infty\\ nh=b-a}} h\,[f(a+h)+f(a+2h)+\ldots+f(a+nh)],$$

$$=\lim_{\substack{h\to 0\\ h\to\infty\\ nh=b-a}} h\,[e^{a+h}+e^{a+2h}+\ldots\ldots e^{a+nh}],$$

$$=\lim_{\substack{h\to 0\\ n\to\infty\\ nh=b\text{ɒ}-}} \frac{he^{a+h}(e^{nh}-1)}{e^h-1},$$

$$=\lim_{h\to 0} \frac{h}{e^h-1}\cdot e^h\cdot e^a\,(e^{b-a}-1),$$

$$=(e^a e^{b-a}-1),\text{ since }\lim_{h\to 0}\frac{h}{e^h-1}=1,\ \lim_{h\to 0} e^h=1.$$

$$=e^b-e^a.$$

Example 3. *Evaluate* $\int_a^b \sin x\,dx$ *by expressing it as the limit of a sum.*

Solution. Here $f(x)=\sin x$.

$$\int_a^b \sin x\,dx$$

$$=\lim_{\substack{h\to 0\\ n\to\infty\\ nh=b-a}} h\,[(\sin(a+h)+\sin(a+2h)+\ldots\ldots+\sin(a+nh)] \quad \cdots(1)$$

Let $S=\sin(a+h)+\sin(a+2h)+\ldots\ldots+\sin(a+nh)]$

Multiplying both sides by $2\sin\frac{1}{2}h$, we have,

$$(2 \sin \tfrac{1}{2}h)\, S = 2 \sin \tfrac{1}{2}h \sin (a+h) + 2 \sin \tfrac{1}{2}h \sin (a+2h) + \cdots\cdots$$
$$+ \cdots\cdots + 2 \sin \tfrac{1}{2}h \sin (a+nh)$$

$$\begin{aligned} &= \cos (a+\tfrac{1}{2}h) - \cos (a+\tfrac{3}{2}h) \\ &+ \cos (a+\tfrac{3}{2}h) - \cos (a+\tfrac{5}{2}h) \\ &\vdots \\ &+ \cos [a+(n-\tfrac{1}{2}h)] - \cos [a+(n+\tfrac{1}{2}h)], \\ &= \cos (a+\tfrac{1}{2}h) - \cos (a+(n+\tfrac{1}{2})h), \\ &= \cos (a+\tfrac{1}{2}h) - \cos (b+\tfrac{1}{2}h). \end{aligned} \qquad \ldots(2)$$

From (1) and (2) we have

$$\int_a^b b \sin x \, dx = \lim_{h \to 0} \frac{h}{2 \sin \frac{1}{2}h} [\cos (b+h) - \cos (a+h)]$$

$$= \cos b - \cos a, \text{ since } \lim_{h \to 0} \frac{\frac{1}{2}h}{\sin \frac{1}{2}h} = 1,$$

$$\lim_{h \to 0} \cos (a+h) = \cos a,$$

$$\lim_{h \to 0} \cos (b+h) = \cos b.$$

EXERCISE 6 (*a*)

Evaluate the following definite integrals by expressing them as limits of sums :

1. $\displaystyle\int_a^b x \, dx.$

2. $\displaystyle\int_a^b (3x+5) \, dx.$

3. $\displaystyle\int_a^b x^2 \, dx.$

4. $\displaystyle\int_a^b x^3 \, dx.$

5. $\displaystyle\int_a^b \cos x \, dx.$

6. $\displaystyle\int_a^b \sin 2x \, dx.$

7. $\displaystyle\int_0^{\pi/2} \sin x \, dx.$

8. $\displaystyle\int_a^b e^{-x} \, dx.$

9. $\displaystyle\int_0^{\pi/8} \cos^2 x \, dx.$

10. $\displaystyle\int_0^1 x e^{x^2} \, dx.$

6·3. THE FUNDAMENTAL THEOREM OF INTEGRAL CALCULUS

Recall that by definition,

$$\int_a^b f(x)\,dx = \lim_{\substack{n\to\infty \\ h\to 0 \\ nh=b-a}} h\,[f(a)+f(a+h)+\cdots\cdots+f(a+\overline{n-1}\ h)]$$

If this has given you the idea that the evaluation of a definite integral is a relatively complicated affair, then we have a pleasant surprise for you. After you are through this section, you would have changed your opinion because you would find that if you can lay your hands on a primitive of $f(x)$ (and you would be able to do this for almost all the functions you are going to come across !),

then evaluating $\int_a^b f(x)\,dx$ will be a very simple affair.

The fundamental theorem of integral calculus, proved by Newton and Leibnitz independently, expresses the definite integral of a function over an interval in terms of the values of a primitive of the function at the end points. The theorem is extremely useful in applications, but its proof is beyond the scope of the present book. We therefore state the theorem without proof.

Theorem 6·1. *(The Fundamental Theorem of Integral Calculus). Let $f(x)$ be a function defined and continuous over a closed interval $[a, b]$. If $F(x)$ is a primitive of $f(x)$, then*

$$\int_a^b f(x)\,dx = F(b)-F(a).$$

The above theorem establishes a relationship between a derivative and an integral. It states that if we are given a function $f(x)$, we can calculate the definite integral of $f(x)$ over $[a, b]$ by finding a primitive of $f(x)$ and then substracting the value of the primitive at a from its value at b. The difference $F(b)-F(a)$ is usually written as

$$F(b)-F(a)=[F(x)]_a^b.$$

Remark. It may be noted that a definite integral has a *definite* value and is independent of the constant of integration. Hence the name *definite* integral. In fact the constant of integration disappears as shown below.

If $F(x)$ is a primitive of $f(x)$, then $F(x)+C$ is also a primitive.

Now, $[F(x)+C]_a^b=$(the value of $F(x)+C$ at $x=b$)$-$(the value of $F(x)+C$ at $x=a$),

$$=(F(b)+C)-(F(a)+C)=F(b)-F(a).$$

Before we proceed to evaluate definite integrals, we prove some simple properties of definite integrals.

We have so far defined $\int_a^b f(x)\,dx$ only when $a<b$. We shall now define $\int_a^b f(x)\,dx$ when $a=b$ and $a>b$. Naturally, we would like our definitions to be such that the fundamental theorem of integral calculus continues to hold.

Definition 6·2. *Let $f(x)$ be defined and have a primitive in $[b, a]$, Then*

(*i*) $\int_a^b f(x)\,dx=0.$

(*ii*) $\int_a^b f(x)\,dx=-\int_b^a f(x)\,dx.$

Remark. Observe that if $F(x)$ be a primitive of $f(x)$, then the fundamental theorem of integral calculus, *if applicable* to $f(x)$ in $[a, b]$, would give

$$\int_a^b f(x)\,dx=\Big[F(x)\Big]_a^a=F(a)-F(a)=0.$$

Also, if the fundamental theorem be applicable to $\int_a^b f(x)\,dx$ where $b<a$, then we would have

$$\int_a^b f(x)\,dx=\Big[F(x)\Big]_a^b$$

$$=F(b)-F(a)=-[F(a)-F(b)]=-\int_b^a f(x)\,dx$$

which is in agreement with the above definition. Thus, with the above definitions we find that the conclusion of the fundamental theorem of integral calculus is valid whether $a<b$, $a=b$, or $a>b$.

Before we proceed to evaluate definite integrals, we prove some simple properties of definite integrals.

Theorem 6·2. *Let* $f(x)$, $g(x)$ *be functions defined and continuous over a closed interval* $[a, b]$. *Then*

$$(i)\quad \int_a^a f(x)\,dx=0.$$

$$(ii)\quad \int_b^a f(x)\,dx=-\int_a^b f(x)\,dx.$$

$$(iii)\quad \int_a^b [f(x)+g(x)]\,dx=\int_a^b f(x)\,dx+\int_a^b g(x)\,dx.$$

$$(iv)\quad \int_a^b k f(x)\,dx=k\int_a^b f(x)\,dx, \text{ where } k \text{ is a constant.}$$

$$(v)\quad \int_a^b f(x)\,dx=\int_a^c f(x)\,dx+\int_c^b f(x)\,dx, \text{ where } c\in[a, b]$$

Proof. (*i*) and (*ii*) are re-statements of definitions.

(*iii*) Suppose F (x) is a primitive of $f(x)$ and G (x) is a primitive of $g(x)$.

Then,

$$\int_a^b f(x)\,dx+\int_a^b g(x)\,dx$$

$$=\Big[\,F(x)\,\Big]_a^b+\Big[\,G(x)\,\Big]_a^b,$$

$$=[F(b)-F(a)]+[G(b)-G(a)],$$

$$=[F(b)+G(b)]-[F(a)+G(a)],$$

$$=\Big[\,F(x)+G(x)\,\Big]_a^b,$$

$$=\int_a^b [f(x)+g(x)]\,dx.$$

[$\because$ F (x)+G (x) is a primitive of $f(x)+g(x)$].

$$(iv)\quad k\int_a^b f(x)\,dx=k\Big[\,F(x)\,\Big]_a^b,$$

$$=k\,[\,F(b)-F(a)],$$

$$=(kF)(b)-(kF)(a),$$

$$=\Big[(kF)(x)\Big]_a^b,$$

$$=\int_a^b kf(x)\,dx.$$

[$\because$ kF (x) is a primitive of $kf(x)$].

$$(v)\quad \int_a^b f(x)\,dx+\int_c^b f(x)\,dx=\Big[\,F(x)\,\Big]_a^c+\Big[\,F(x)\,\Big]_c^b,$$

$$=(F(c)-F(a))+(F(b)-F(c))$$

$$=F(b)-F(a),$$

$$=\Big[\,F(x)\,\Big]_a^b,$$

$$=\int_a^b f(x)\,dx.$$

6·4. EVALUATION OF DEFINITE INTEGRALS

The fundamental theorem of integral calculus shows that to evaluate the definite integral of a function, we need only to find its primitive. Various methods for determining primitives of different types of functions developed in Chapter 5 will now be used to calculate definite integrals.

Example 4. *Evaluate* $\int_0^1 (x^2+5x)\,dx.$

Solution. $\int_0^1 (x^2+5x)\,dx=\int_0^1 x^2\,dx+\int_0^1 5x\,dx,$

$$=\left[\frac{x^3}{3}\right]_0^1+\left[\frac{5x^3}{2}\right]_0^1$$

$$=\left(\frac{1}{3}-0\right)+\left(\frac{5}{2}-0\right)$$

$$=\frac{1}{3}+\frac{5}{2}=\frac{17}{6}.$$

Example 5. *Evaluate* $\int_0^{\pi} x \sin x \, dx.$

Solution.

Integrating by parts, we get

$$\int x \sin x \, dx = x(-\cos x) - \int 1(-\cos x) dx,$$

$$= -x \cos x + \sin x + C.$$

Thus $-x \cos x + \sin x$ is a primitive of $x \sin x$.

$$\therefore \int_0^{\pi} x \sin x \, dx = \Big[-x \cos x + \sin x\Big]_0^{\pi}$$

$$= \pi.$$

6·4·1. Evaluation of the definite integral

$$\int_0^{\pi/2} \sin^m x \cos^n x \, dx,$$

where m, n are any positive integers.

$$\text{Let } I_{m,n} = \int_0^{\pi/2} \sin^m x \cos^n x \, dx.$$

From the reduction formula (*a*) of Example 51, Chapter 5, we have

$$(m+n)I_{m,n} = \Big[-\sin^{m-1}x \cos^{n+1} x\Big]_0^{\pi/2} + (m-1).\ I_{m-2,n}.$$

$$=(m-1)I_{m-2,n},$$

or $$I_{m,n}=\frac{m-1}{m+1}\,I_{m-2,n}.$$

Changing m to $m-2$, we have

$$I_{m-2,n}=\frac{m-3}{m+n-2}\,I_{m-4,n},$$

$$I_{m-4,n}=\frac{m-5}{m+n-4}\,I_{m-6,\,n},$$

....

$$I_{3,n}=\frac{2}{3+n}\cdot I_{1,n},\text{ if } m \text{ is odd}$$

$$I_{2,n}=\frac{1}{2+n}\cdot I_{0,n},\text{ if } m \text{ is even}$$

Thus

$$I_{m,n}=\begin{cases}\dfrac{m-1}{m+1}\cdot\dfrac{m-3}{m+n-2}\cdot\dfrac{m-5}{m+n-4}\cdot\;\dfrac{2}{3+n}\cdot I_{1,n}, & \\ \qquad\qquad\qquad\text{if } m \text{ is odd} & \ldots(1)\\ \dfrac{m-1}{m+n}\cdot\dfrac{m-3}{m+n-4}\cdot\dfrac{m-5}{m+n-4}\cdot\;\dfrac{1}{2+n}\cdot I_{0,n}, & \\ \qquad\qquad\qquad\text{if } m \text{ is even} & \ldots(2)\end{cases}$$

Now

$$I_{1,n}=\int_0^{\pi/2}\sin x\cos^n x\,dx=\left[\frac{-\cos^{n+1}x}{n+1}\right]_0^{\pi/2}=\frac{1}{n+1}\qquad\ldots(3)$$

$$I_{0,n}=\int_0^{\pi/2}\sin^0 x\cos^n x\,dx=\int_0^{\pi/2}\cos^n x\,dx\qquad\ldots(4)$$

From the reduction formulae (b) of Example 51, Chapter 5, we have on putting $m=0$,

$$nI_{0,n}=\Big[\sin x\;\cos^{n-1}x\Big]_0^{\pi/2}+(n-1)I_{0,n-2},$$

or $$I_{0,n}=\frac{n-1}{n}\,I_{0,\,n-2}.$$

Changing n to $n-2$, we have

$$I_{0,n-2}=\frac{n-3}{n-2}\,I_{0,\,n-4},$$

$$I_{0,n-4}=\frac{n-5}{n-4}\,I_{0,\,n-6},$$

....

....

$$I_{0,3}=\frac{2}{3}\,I_{0,1}\text{ if } n \text{ is odd}$$

$$I_{0,\,2}=\frac{1}{2}\,I_{0,\,0}\text{ if } n \text{ is even}$$

Thus

$$I_{0,\,n}=\begin{cases}\dfrac{n-1}{n}\cdot\dfrac{n-3}{n-2}\cdot\dfrac{n-5}{n-4}\cdot\cdots\cdots\dfrac{2}{3}\,I_{0,\,1}\text{ if } n \text{ is odd} & \dots(5)\\[2ex] \dfrac{n-1}{n}\cdot\dfrac{n-3}{n-2}\cdot\dfrac{n-5}{n-4}\cdots\cdots\dfrac{1}{2}\,I_{0,\,0}\text{ is } n \text{ is even} & \dots(6)\end{cases}$$

$$I_{0,1}=\int_0^{\pi/2}\sin^0 x\ \cos^1 x\,dx=\Big[\sin x\Big]_0^{\pi/2}=1 \qquad \dots(7)$$

$$I_{0,0}=\int_0^{\pi/2}\sin^0 x\cos^0 x\,dx=\Big[\ x\ \Big]_0^{\pi/2}=\pi/2. \qquad \dots(8)$$

From (1) and (3) we find that *if m is odd,*

$$I_{m,n}=\frac{m-1}{m+n}\cdot\frac{m-3}{m+n-2}\cdot\frac{m-5}{m+n-4}\cdots\cdots\frac{2}{3+n}\cdot\frac{1}{1+n}, \qquad \dots(9)$$

whether n is odd or even.

From (2), (4), (5) and (7) we find that if m is even and n is odd then

$$I_{m,n}=\frac{m-1}{m+n}\cdot\frac{m-3}{m+n-2}\cdot\frac{m-5}{m+n-2}\cdots\cdots\frac{1}{2+n}\cdot\frac{n-1}{n}\cdot\frac{n-3}{n-2}\cdots\cdots\frac{2}{3} \quad \dots(10)$$

From (2), (4), (6) and (8) we find that if m is even and n is also even, then

$$I_{m,n}=\frac{m-1}{m+n}\cdot\frac{m-3}{m+n-2}\cdots\cdots\frac{1}{2+n}\cdot\frac{n-1}{n}\cdot\frac{n-3}{n-2}\cdots\cdots\frac{1}{2}\cdot\frac{\pi}{2} \qquad \dots(11)$$

The above formulae (9), (10) and (11) can be combined to have the following simple formulae covering all the cases :

$$\int_0^{\pi/2} \sin^m x \cos^n x \; dx = \frac{(m-1)(m-3)\ldots\ldots(n-1)(n-3)\ldots\ldots}{(m+n)(m+n+2)\ldots\ldots\ldots\ldots} \times K,$$

the three sets of factors starting with $m-1$, $n-1$ and $m+n$ and diminishing by 2 at a time, end up with either 1 or 2 according as the first factor of the set is odd or even, and $K=\pi/2$ if m and n are both even, and $K=1$ if at least one of m and n is odd.

Example 6.

Write down the value of

(*a*) $\int_0^{\pi/2} \sin^5 x \cos^7 x \, dx$ (*b*) $\int_0^{\pi/2} \sin^6 x \cos^8 x \, dx.$

Solution.

$$I_{5,7} = \int_0^{\pi/2} \sin^5 x \cos^7 x \, dx = \frac{4.2.6.4.2}{12.10.8.6.4.2} \times K$$

$$= \frac{k}{120}, \text{ where K is 1 since the exponents are } not \text{ both even.}$$

$$I_{6,8} = \int_0^{\pi/2} \sin^6 x \cos^8 x \, dx = \frac{5.3.1.7.5.3.1}{14.12.10.8.6.4.2} K$$

$$= \frac{5K}{2048}, \text{ where K is } \frac{\pi}{2} \text{ since the exponents are both even.}$$

$$= \frac{5\pi}{4096}.$$

EXERCISE 6 (*b*)

Evaluate the following definite integrals :

1. $\int_0^4 x^3 \, dx.$ 2. $\int_0^3 (2x^2+5x) \, dx.$

3. $\int_0^1 \frac{2x}{1+x^2} \, dx.$ 4. $\int_0^1 e^x \, dx.$

5. $\int_0^{\pi/6} \sec^2 x\, dx.$

6. $\int_2^3 \frac{(x^2+3)^2}{x^2}\, dx.$

7. $\int_0^{\frac{1}{2}} \frac{dx}{\sqrt{1-x^2}}.$

8. $\int_0^1 \frac{x\, dx}{\sqrt{(1+x^2)}}.$

9. $\int_0^{\pi/4} \tan x\, dx.$

10. $\int_0^2 \frac{dx}{x^2+4}.$

11. $\int_0^1 x \sin^{-1} x\, dx.$

12. $\int_0^1 x \tan^{-1} x\, dx.$

13. $\int_0^{\pi} x \cos x\, dx.$

14. $\int_0^{\pi/4} x \sin 2x\, dx.$

15. $\int_0^1 x^2 e^{2x}\, dx.$

16. $\int_0^1 \frac{dx}{x^2+x+1}.$

17. $\int_0^{\pi/2} \sin^3 t\, dt.$

18. $\int_0^{\pi/2} \cos^4 x\, dx.$

19. $\int_0^{\pi/2} \sin^2 x \cos^4 x\, dx.$

20. $\int_0^{\pi/2} \sin^3 x \cos^5 x\, dx.$

6·5. TRANSFORMATION OF DEFINITE INTEGRALS BY SUBSTITUTION

In the preceding chapter we have seen that the method of substitution is a powerful tool for finding primitives. In view of the fundamental theorem of integral calculus, we should naturally expect that the method of substitution would prove useful for evaluating definite integrals. This is indeed the case. Evaluation of a definite integral $\int_a^b f(x)\, dx$ by employing the fundamental theorem of integral calculus and the method of substitution consists of the following steps :

1. Find a suitable substitution $x=\phi(t)$ which is going to be useful in obtaining the value of

$$\int f(x)\,dx.$$

2. Use the substitution $x=\phi(t)$ to transform

$$\int f(x)\,dx \text{ into} \int f(\phi(t))\,\phi'(t)\,dt.$$

3. Evaluate $\int f(\phi(t))\phi'(t)\,dt$ to get the primitive $F(t)+C$.

4. Transform the primitive by using the inverse of $x=\phi(t)$, say $t=\psi(x)$ to get $F(\psi(x))+C$.

5. Use the fundamental theorem of integral calculus to get the value of the integral as

$$F(\psi(b))-F(\psi(a)).$$

Out of the five steps listed above, step 4 is often not easily manageable, and we use a different procedure, namely, the limits of integration are changed. We try to find the values of t when $x=a$ and $x=b$, and transform the *definite* integral as such rather than transforming only the *indefinite* integral. This is done by means of the following theorem :

Theorem 6·3. (*Substitution Rule*)

Let $f(x)$ be a function which has a primitive. Let $x=\phi(t)$ be a function and α, β be real numbers such that $\phi(\alpha)=a$, $\phi(\beta)=b$. If $\phi(t)$ possesses a continuous derivative on $[\alpha, \beta]$, then

$$\int_a^b f(x)\,dx=\int_\alpha^\beta f(\phi(t))\,\phi'(t)\,dt.$$

Proof. Let $F(x)$ be a primitive of $f(x)$, so that $F'(x)=f(x)$. By the fundamental theorem of integral calculus,

$$\int_a^b f(x)\,dx=F(b)-F(a)$$

$$=F(\phi(\beta))-F(\phi(\alpha)) \qquad \ldots(1)$$

Let us put $G(t)=F(\phi(t))$. By the chain rule for differentiation, $G'(t)=F'(\phi(t))\,\phi'(t)=f(\phi(t))\,\phi'(t)$, so that $G(t)$ is a primitive of $f(\phi(t))\,\phi'(t)$, *i.e.*,

$$\int_\alpha^\beta f(\phi(t))\,\phi'(t)\,dt=G(\beta)-G(\alpha)=F(\phi(\beta))-F(\phi(\alpha)) \quad \ldots(2)$$

By comparing (1) and (2), we have

$$\int_a^b f(x)\,dx=\int_\alpha^\beta f(\phi(t))\,\phi'(t)\,dt.$$

Remark. By interchanging the roles of x and t we can write the above in the following form :

$$\int_\alpha^\beta f(\phi(x))\,\phi'(x)\,dx=\int_a^b f(t)\,dt.$$

Example 7. *Evaluate* :

$$\int_0^{\pi/2} \frac{dx}{1+2\sin x+\cos x}.$$

Solution.

$$\text{Let } I=\int_0^{\pi/2} \frac{dx}{1+2\sin x+\cos x}.$$

Put $\tan(x/2)=t$, so that $\frac{1}{2}\sec^2(x/2)\,dx=dt$, *i.e.*, $dx=\dfrac{2dt}{1+t^2}$

When $x=0$, $t=0$; when $x=\dfrac{\pi}{2}$, $t=1$.

Therefore

$$I=\int_0^1 \frac{1}{1+2\left(\frac{2t}{1+t^2}\right)+\frac{1-t^2}{1+t^2}}\cdot\frac{2dt}{1+t^2},$$

$$=\int_0^1 \frac{dt}{2t+1},$$

$$=\left[\frac{1}{2}\,ln\,|\,2t+1\,|\,\right]_0^1,$$

$$=\frac{1}{2}\,ln\,3.$$

Example 8. *Evaluate* :

(*a*) $$\int_1^2 \frac{x^2+1}{x^4+1}\,dx.$$

(b) $\int_1^2 \frac{x^2-1}{x^4+1}\,dx.$

Solution.

(a) Let $I=\int_1^2 \frac{x^2+1}{x^4+1}\,dx,$

$$=\int_1^2 \frac{1+\frac{1}{x^2}}{x^2+\frac{1}{x^2}}\,dx.$$

Put $x-\frac{1}{x}=t$, so that $\left(1+\frac{1}{x^2}\right)dx=dt.$

Also, when $x=1$, $t=0$, and when $x=2$, $t=3/2$.

$$\therefore \qquad I=\int_0^{3/2} \frac{dt}{t^2+2}=\frac{1}{\sqrt{2}}\left[\tan^{-1}\ \frac{t}{\sqrt{2}}\right]_0^{3/2},$$

$$=\frac{1}{\sqrt{2}}\tan^{-1}\left(\frac{3}{2\sqrt{2}}\right).$$

(b) Let $I=\int_1^2 \frac{x^2-1}{x^4+1}\,dx,$

$$=\int_1^2 \frac{1-\frac{1}{x^2}}{x^2+\frac{1}{x^2}}\,dx.$$

Put $x+\frac{1}{x}=t$, so that $\left(1-\frac{1}{x^2}\right)dx=dt.$

Also, when $x=1$, $t=0$; when $x=2$, $t=\frac{3}{2}.$

$$\therefore \qquad I=\int_0^{3/2} \frac{dt}{t^2-2}=\left[\frac{1}{2\sqrt{2}}\ ln\ \left|\frac{t-\sqrt{2}}{t+\sqrt{2}}\right|\right]_0^{3/2}$$

$$=\left(\frac{1}{\sqrt{2}}\right) ln\,(3-2\sqrt{2}).$$

Example 9. *Evaluate* :

$$\int_{\alpha}^{\beta} (x-\alpha)(\beta-x)\, dx \qquad (0<\alpha<\beta)$$

Solution. Let $x=\alpha \cos^2 \theta+\beta \sin^2 \theta$,

so that $dx=2(\beta-\alpha) \sin \theta \cos \theta \, d\theta$.

$$x-\alpha=(\beta-\alpha) \sin^2 \theta,\ \beta-x=(\beta-\alpha) \cos^2 \theta,$$

so that, $\sqrt{(x-\alpha)(\beta-x)}=(\beta-\alpha) \sin \theta \cos \theta$.

Also, when $x=\alpha$, $\theta=0$; when $x=\beta$, $\theta=\frac{\pi}{2}$.

$$\therefore \quad I=\int_{\alpha}^{\beta} \sqrt{(x-\alpha)(\beta-x)}\, dx,$$

$$=\int_{0}^{\pi/2} (\beta-\alpha)^2 \sin^2 \theta \cos^2 \theta \, d\theta,$$

$$=(\beta-\alpha)^2 \int_{0}^{\pi/2} \sin^2 \theta \cos^2 \theta \, d\theta,$$

$$=(\beta-\alpha)^2 \frac{1.1}{4.2} \cdot \frac{\pi}{2},$$

$$=\frac{\pi}{16} (\beta-\alpha)^2.$$

EXERCISE 6 (*c*)

Evaluate :

1. $\int_{0}^{\pi/2} (1+\sin x)^2 \cos x \, dx$.
2. $\int_{0}^{1} \frac{x^2}{1+x^3}\, dx$.
3. $\int_{0}^{1} x\sqrt{x^2+4}\, dx$.
4. $\int_{0}^{\pi/2} \sqrt{\sin x} \cos x \, dx$.
5. $\int_{0}^{9} \frac{dx}{1+\sqrt{x}}$.
6. $\int_{0}^{1} \frac{dx}{e^x+e^{-x}}$.

7. $\int_1^3 \frac{\cos(\ln x)}{x}\,dx.$

8. $\int_3^6 \frac{dx}{x(x^4-1)}.$

9. $\int_{\frac{1}{4}}^{\frac{1}{2}} \frac{dx}{\sqrt{(x-x^2)}}.$

10. $\int_0^1 \frac{\sqrt{(\tan^{-1} x)}}{1+x^2}\,dx.$

11. $\int_1^2 \frac{dx}{x(x^3+1)}$

12. $\int_0^{\pi/4} \frac{\cos x}{\sqrt{1-2\sin^2 x}}\,dx.$

13. $\int_0^1 x(\tan^{-1} x)^2\,dx.$

14. $\int_0^6 \frac{x\,dx}{(x^2+4)(x^2+9)}.$

15. $\int_0^1 \sin^{-1}\sqrt{\frac{x}{1+x}}\,dx.$

16. $\int_0^{1/\sqrt{2}} \frac{\sin^{-1} x}{(1-x^2)^{3/2}}\,dx.$

17. $\int_0^{\pi/2} \frac{dx}{4+3\cos x}.$

18. $\int_0^{\pi/2} \frac{dx}{4+5\sin x}.$

19. $\int_0^{\pi} \frac{dx}{1-2a\cos x+a^2}.$

20. $\int_0^{\pi/2} \frac{\cos x\,dx}{(1+\sin x)(2+\sin x)}.$

21. $\int_0^{\pi} \frac{dx}{3+2\sin x+\cos x}.$

6·6. USE OF DEFINITE INTEGRALS TO FIND THE SUM OF A SERIES

We have already seen that if f is continuous on $[a, b]$, then

(A) $\lim_{n\to\infty} [h\{f(a)+f(a+h)+\cdots+f(a+\overline{n-1}\,h)\}],$

$$=\int_a^b f(x)\,dx,$$

where $nh=b-a$

and

(B) $\lim\limits_{n\to\infty} [h\{f(a+h)+f(a+2h)+\cdots+f(a+nh)\}]$,

$$=\int_a^b f(x)dx,$$

where $\quad nh=b-a$

The above formulae are often used to express the sum of a series as a definite integral. Whenever we can put a series in the form given in the left hand side of (A) or (B), then the sum of the series equals $\int_a^b f(x)dx$. By evaluating this integral we can obtain the sum of the series. The special cases of (A) and (B) when $a=0$, $b=1$ are of interest and therefore we state them below for ready reference :

(C) $\lim\limits_{\substack{n\to\infty\\ nh=1}} [h\{f(0)+f(h)+f(2h)+\cdots+f(\overline{n-1}h)\}]=\int_0^1 f(x)dx.$

(D) $\lim\limits_{\substack{n\to\infty\\ nh=1}} [h\{f(h)+f(2h)+f(3h)+\cdots+f(nh)\}]=\int_0^1 f(x)dx.$

The following examples illustrate as to how the formulae (A) and (D) are used.

Example 10. *Find the limit as $n\to\infty$ of the sum*

$$\frac{1}{n}+\frac{n}{(n+1)^2}+\frac{n}{(n+2)^2}+\cdots+\frac{n}{(2n-1)^2}$$

Solution.

Let $S_n=\dfrac{1}{n}+\dfrac{n}{(n+1)^2}+\dfrac{n}{(n+2)^2}+\cdots+\dfrac{n}{(2n-1)^2}$,

$$=\frac{1}{n}\left[1+\frac{1}{\left(1+\frac{1}{n}\right)^2}+\frac{1}{\left(1+\frac{2}{n}\right)^2}+\cdots\frac{1}{\left(1+\frac{n-1}{n}\right)^2}\right],$$

$$=h\left[1+\frac{1}{(1+h)^2}+\frac{1}{(1+2h)^2}+\cdots\frac{1}{(1+\overline{n-1}h)^2}\right],$$

where $\quad nh=1$

$$=h\{f(0)+f(h)+f(2h)+\cdots+f(\overline{n-1}\,h)\},$$

where $\quad f(x)=\dfrac{1}{(1+x)^2}$, $nh=1$.

Since $\quad nh=1$, therefore when $n\to\infty$, $h\to 0$,

$$\therefore \quad \lim_{n\to\infty} S_n = \lim_{\substack{n\to\infty \\ h\to 0 \\ nh=1}} h[(f(0)+f(h)+f(2h)+\dots+f(\overline{n-1}\,h)]$$

$$=\int_0^1 f(x)dx$$

$$=\int_0^1 \frac{dx}{(1+x)^2}.$$

$$=\left[-\frac{1}{1+x}\right]_0^1=\frac{1}{2}.$$

Remark. Here we have applied formula (*C*).

Example 11. *Find the limit, when n tends to infinity of*

$$\frac{1}{\sqrt{(2n-1^2)}}+\frac{1}{\sqrt{(4n-2^2)}}+\frac{1}{\sqrt{(6n-3^2)}}+\dots+\frac{1}{n}.$$

Solution.

$$\text{Let } S_n=\frac{1}{\sqrt{(2n-1^2)}}+\frac{1}{\sqrt{(4n-2^2)}}+\frac{1}{\sqrt{(6n-3^2)}}+\dots+\frac{1}{n}$$

$$=\frac{1}{n}\left\{\frac{n}{\sqrt{(2n-1^2)}}+\frac{n}{\sqrt{(4n-2^2)}}+\frac{n}{\sqrt{6n-3^2)}}+\dots+\frac{n}{n}\right\}$$

$$=\frac{1}{n}\left\{\frac{1}{\sqrt{\left(\frac{2}{n}-\left(\frac{1}{n}\right)^2\right)}}+\frac{1}{\sqrt{\left(\frac{4}{n}-\left(\frac{2}{n}\right)^2\right)}}+\dots+\frac{1}{\sqrt{\left(\frac{2n}{n}-\left(\frac{n}{n}\right)^2\right)}}\right\},$$

$$=h\left\{\frac{1}{\sqrt{(2h-h^2)}}+\frac{1}{\sqrt{(4h-(2h)^2)}}+\dots+\frac{1}{\sqrt{(2nh-(nh)^2)}}\right\}$$

where $\quad nh=1$

$$=h\{f(h)+f(2h)+\dots+f(nh)\},$$

where $nh=1$, $f(x)=\dfrac{1}{\sqrt{(2x-x^2)}}$

Since $\quad nh=1$, therefore $h\to 0$ as $n\to\infty$.

$$\therefore \quad \lim_{n\to\infty} S_n = \lim_{\substack{n\to\infty \\ h\to 0 \\ nh=1}} h\{(f(h)+f(2h)+\ldots+f(nh)\},$$

$$=\int_0^1 f(x)dx,$$

$$=\int_0^1 \frac{dx}{\sqrt{(2x-x^2)}},$$

$$=\int_0^1 \frac{1}{\sqrt{\{1-(1-x)^2\}}},$$

$$=\Big[-\sin^{-1}(1-x)\Big]_0^1=\pi/2$$

Remark. Here we have applied formula (D).

Example 12. *Find the limit, when n tends to infinity of*

$$\frac{1}{n}+\frac{1}{n+1}+\frac{1}{n+2}+\ldots+\frac{1}{4n}.$$

Solution.

First of all let us observe that since each term tends to zero, and there are $3n+1$ terms in all, the limit will remain the same if the first term is left out.

Let us write

$$S_n=\frac{1}{n+1}+\frac{1}{n+2}+\ldots+\frac{1}{4n}$$

$$=\left(\frac{1}{n+1}+\frac{1}{n+2}+\ldots+\frac{1}{2n}\right)$$

$$+\left(\frac{1}{2n+1}+\frac{1}{2n+2}+\ldots+\frac{1}{3n}\right)$$

$$+\left(\frac{1}{3n+1}+\frac{1}{3n+2}+\ldots+\frac{1}{4n}\right),$$

$$=A_n+B_n+C_n,$$

where

$$A_n=\frac{1}{n+1}+\frac{1}{n+2}+\ldots+\frac{1}{2n},$$

$$B_n=\frac{1}{2n+1}+\frac{1}{2n+2}+\ldots+\frac{1}{3n},$$

$$C_n=\frac{1}{2n+1}+\frac{1}{2n+1}+\dots+\frac{1}{3n}.$$

To find limits of A_n, B_n, C_n as n tends to infinity, write

$$D_n=\frac{1}{kn+1}+\frac{1}{kn+2}+\dots+\frac{1}{kn+n},$$

and observe that $D_n=A_n$, B_n or C_n according as $k=1$, 2 or 3.

Now $$D_n=\frac{1}{n}\left\{\frac{1}{k+h}+\frac{1}{k+2h}+\dots+\frac{1}{k+nh}\right\}, \text{where } nh=1$$

$$=h\{f(h)+f(2h)+\dots+f(nh)\},$$

where $f(x)=\dfrac{1}{k+x}$, $nh=1$.

$$\therefore \lim_{n\to\infty} D_n=\int_0^1 f(x)\,dx=\int_0^1\frac{1}{k+x}\,dx=\Big|\ln|k+x|\Big|_0^1$$

$$=\ln\left|\frac{k+1}{k}\right|.$$

Putting $k=1$, 2, 3 in succession we have

$$\therefore \lim_{n\to\infty} A_n=\ln 2,\ \lim_{n\to\infty} B_n=\ln\left(\frac{3}{2}\right),\ \lim_{n\to\infty} C_n=\ln\left(\frac{4}{3}\right)$$

$$\therefore \lim_{n\to\infty} S_n=\lim_{n\to\infty}(A_n+B_n+C_n)$$

$$=\lim_{n\to\infty} A_n+\lim_{n\to\infty} B_n+\lim_{n\to\infty} C_n$$

$$=\ln 2+\ln\frac{3}{2}+\ln\frac{4}{3},$$

$$=\ln 4.$$

EXERCISE 6 (*d*)

Find the limit, when n tends to infinity, of the following :

1. $\dfrac{1}{n+1}+\dfrac{1}{n+2}+\dfrac{1}{n+3}+\dots+\dfrac{1}{n+n}$.

2. $\dfrac{1}{n}+\dfrac{n^2}{(n+1)^3}+\dfrac{n^2}{(n+2)^3}+\dfrac{n^2}{(n+3)^3}+\dots+\dfrac{1}{8n}$.

3. $\dfrac{n+1}{n^2+1^2}+\dfrac{n+2}{n^2+2^2}+\dfrac{n+3}{n^2+3^2}+\dots+\dfrac{1}{n}$.

4. $\dfrac{n}{n^2+1^2}+\dfrac{n}{n^2+2^2}+\dfrac{n}{n^2+3^2}+\dots+\dfrac{1}{2n}.$

5. $\dfrac{1}{n}+\dfrac{1}{n+1}+\dfrac{1}{n+2}+\dots+\dfrac{1}{3n}.$

6. $\dfrac{1}{\sqrt{n^2}}+\dfrac{1}{\sqrt{(n^2-1)}}+\dfrac{1}{\sqrt{(n^2-2^2)}}+\dots+\dfrac{1}{\sqrt{[n^2-(n-1)^2]}}.$

7. $\dfrac{1}{n^3}+\dfrac{4}{n^3}+\dfrac{9}{n^3}+\dots+\dfrac{1}{n}.$

8. $\dfrac{1}{1^3+n^3}+\dfrac{2^2}{2^3+n^3}+\dfrac{3^2}{3^3+n^3}+\dots+\dfrac{n^2}{n^3+n^3}.$

9. $\dfrac{\sqrt{n}}{\sqrt{n^3}}+\dfrac{\sqrt{n}}{\sqrt{(n+4)^3}}+\dfrac{\sqrt{n}}{\sqrt{(n+8)^3}}+\dots+\dfrac{\sqrt{n}}{\sqrt{[n+4(n-1)]^3}}.$

10. $\dfrac{n^2}{(n^2+1)^{3/2}}+\dfrac{n^2}{(n^2+2^2)^{3/2}}+\dfrac{n^2}{(n^2+3^2)^{3/2}}+\dots$

$$+\frac{n^2}{[n^2+(n-1)^2]^{3/2}}.$$

11. $\dfrac{\sqrt{n}}{(3+4\sqrt{n})^2}+\dfrac{\sqrt{n}}{\sqrt{2}\,(3\sqrt{2}+4\sqrt{n})^2}+\dfrac{\sqrt{n}}{\sqrt{3}\,(3\sqrt{3}+\sqrt{n})^2}$

$$+\dots+\frac{1}{49\,n}.$$

12. $\displaystyle\sum_{r=1}^{n}\frac{r^3}{r^4+n^4}.$ 13. $\displaystyle\sum_{r=0}^{n-1}\frac{\sqrt{n^2-r^2}}{n^2}.$

14. $\displaystyle\sum_{r=0}^{n-1}\frac{1}{n}\sqrt{\left(\frac{n+r}{n-r}\right)}.$ 15. $\displaystyle\sum_{r=1}^{3n}\frac{n^2}{(3n+r)^2}.$

6·7. SOME MORE PROPERTIES OF DEFINITE INTEGRALS

You have already studied some properties of definite integrals in Section 6·2. We now give some more properties of definite integrals which are found to be useful in evaluating definite integrals. We shall also illustrate their use in evaluating definite integrals.

Theorem 6·4. *If $f(x)$ has a primitive, then*

(i) $\displaystyle\int_0^a f(x)\,dx=\int_0^a f(a-x)dx.$

(*ii*) $\int_0^{2a} f(x)\,dx=\int_0^a f(x)\,dx+\int_0^a f(2a-x)\,dx.$

Proof.

(*i*) Put $x=a-t$, so that $dx=-dt$.

Also, $t=a$ when $x=0$, and $t=0$ when $x=a$.

$$\therefore \int_0^a f(x)\,dx=-\int_a^0 f(a-t)\,dt,$$

$$=\int_0^a f(a-t)\,dt,$$

$$=\int_0^a f(a-x)\,dx.$$

(*ii*) Since $0<a<2a$, therefore

$$\int_0^{2a} f(x)\,dx=\int_0^a f(x)\,dx+\int_a^{2a} f(x)\,dx \qquad ...(1)$$

In the second integral on the right
put $x=2a-t$, so that $dx=-dt$.
Also $t=a$ when $x=a$, and $t=0$ when $x=2a$.

$$\therefore \int_a^{2a} f(x)\,dx=-\int_a^0 f(2a-t)\,dt=\int_0^a f(2a-t)\,dt=\int_0^a f(2a-x)\,dx. \qquad ...(2)$$

From (1) and (2), we have

$$\int_0^{2a} f(x)\,dx=\int_0^a f(x)\,dx+\int_0^a f(2a-x)\,dx.$$

Corollary 1. If $f(2a-x)=f(x)$, then (2) gives

$$\int_0^{2a} f(x)\,dx=\int_0^a f(x)\,dx+\int_0^a f(x)\,dx=2\int_0^a f(x)\,dx.$$

Corollary 2. If $f(2a-x)=-f(x)$, then (2) gives

$$\int_0^{2a} f(x)\,dx=\int_0^a f(x)\,dx+\int_0^a -f(x)\,dx=\int_0^a f(x)\,dx-\int_0^a f(x)\,dx=0.$$

The above corollaries are useful in evaluating

$$\int_0^{\pi} \sin^m x \cos^n x\,dx \text{ and } \int_0^{2\pi} \sin^m x \cos^n x\,dx.$$

(A) Let $I=\int_0^{\pi} \sin^m x \cos^n x\,dx.$

Writing $f(x)=\sin^m x \cos^n x\,dx$, we find that

$$\begin{aligned} f(\pi-x)&=\sin^m(\pi-x)\cos^n(\pi-x),\\ &=\sin^m x\,[-\cos x]^n,\\ &=(-1)^n \sin^m x \cos^n x,\\ &=(-1)^n f(x). \end{aligned}$$

$$\therefore \quad f(\pi-x)=\begin{cases} f(x), \text{ if } n \text{ is even}\\ -f(x), \text{ if } n \text{ is odd}\end{cases}$$

By Corollary 1, we find that

$$I=\int_0^{\pi} f(x)\,dx=2\int_0^{\pi/2} f(x)\,dx, \text{ if } n \text{ is even}$$

By Corollary 2, we find that

$$I=\int_0^{\pi} f(x)\,dx=0, \text{ if } n \text{ is odd.}$$

Thus

$$\int_0^{\pi} \sin^m x \cos^n x\,dx=\begin{cases} 2\int_0^{\pi/2} \sin^m x \cos^n x\,dx, \text{ if } n \text{ is even}\\ 0, \text{ if } n \text{ is odd}\end{cases} \quad \ldots(3)$$

(B) Let $J=\int_0^{2\pi} \sin^m x \cos^n x\,dx.$

Writing $f(x)=\sin^m x \cos^n x$, we find that

$$f(2\pi-x)=\sin^m(2\pi-x)\cos^n(2\pi-x),$$
$$=(-1)^m \sin^m x \cos^n x,$$
$$=(-1)^m f(x).$$

$$\therefore \quad f(2\pi-x)=\begin{cases} f(x), & \text{if } m \text{ is even} \\ -f(x), & \text{if } m \text{ is odd} \end{cases}$$

By Corrollary 1, we find that

$$\int_0^{2\pi} f(x)\,dx=\begin{cases} 2\int_0^{\pi} f(x)\,dx, & \text{if } m \text{ is even} \\ 0, & \text{if } m \text{ is odd} \end{cases} \quad \text{...(4)}$$

By (3),
$$\int_0^{\pi} f(x)\,dx=\begin{cases} 2\int_0^{\pi/2} f(x)\,dx, & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd} \end{cases} \quad \text{...(5)}$$

From (4) and (5) we find that

if m and n are both even, then

$$\int_0^{2\pi} \sin^m x \cos^n x\,dx=4\int_0^{\pi/2} \sin^m x \cos^n x\,dx,$$

and if either m is odd or n is odd, then

$$\int_0^{2\pi} \sin^m x \cos^n x\,dx=0.$$

Example 13. *Evaluate*

(*a*) $\int_0^{\pi} \sin^4 x \cos^2 x\,dx$ (*b*) $\int_0^{\pi} \sin^4 x \cos^3 x\,dx$.

Solution.

(*a*) Since $\sin^4(\pi-x)\cos^2(\pi-x)=\sin^4 x\cos^2 x$,

$$\text{therefore} \int_0^{\pi} \sin^4 x \cos^2 x\,dx=2\int_0^{\pi/2} \sin^4 x \cos^2 x\,dx$$

$$=2.\ \frac{3.1.1}{6.4.2}\cdot\frac{\pi}{2}=\frac{\pi}{16}.$$

(*b*) Since $\sin^4(\pi-x)\cos^3(\pi-x)=-\sin^4 x\cos^3 x$,

$$\therefore \quad \int_0^{\pi} \sin^4 x \cos^3 x \, dx = 0.$$

Example 14. *Evaluate :*

(*a*) $\int_0^{2\pi} \sin^3 x \cos^4 x \, dx$ (*b*) $\int_0^{2\pi} \sin^4 x \cos^3 x \, dx$

(*c*) $\int_0^{2\pi} \sin^4 x \cos^6 x \, dx.$

Solution.

(*a*) Since $\sin^3 (2\pi - x) \cos^4 (2\pi - x) = -\sin^3 x \cos^4 x$,

$$\therefore \quad \int_0^{2\pi} \sin^3 x \cos^4 x \, dx = 0.$$

(*b*) Since $\sin^4 (2\pi - x) \cos^3 (2\pi - x) = \sin^4 x \cos^3 x$,

$$\therefore \quad \int_0^{2\pi} \sin^4 x \cos^3 x \, dx = 2 \int_0^{\pi} \sin^4 x \cos^3 x \, dx \quad \text{...(1)}$$

Again, since

$$\sin^4 (\pi - x) \cos^3 (\pi - x) = -\sin^4 x \cos^3 x,$$

$$\therefore \quad \int_0^{\pi} \sin^4 x \cos^3 x \, dx = 0. \quad \text{...(2)}$$

From (1) and (2), we find that

$$\int_0^{2\pi} \sin^4 x \cos^3 x \, dx = 0.$$

(*c*) Since $\sin^4 (2\pi - x) \cos^6 (2\pi - x) = \sin^4 x \cos^6 x$,

$$\therefore \int_0^{2\pi} \sin^4 x \cos^6 x \, dx = 2 \int_0^{\pi} \sin^4 x \cos^6 x \, dx. \quad \text{...(3)}$$

Again, since $\sin^4 (\pi - x) \cos^6 (\pi - x) = \sin^4 x \cos^6 x$,

$$\text{therefore} \int_0^{\pi} \sin^4 x \cos^6 x \, dx = 2 \int_0^{\pi/2} \sin^4 x \cos^6 x \, dx \quad \text{...(4)}$$

Also $\int_0^{\pi/2} \sin^4 x \cos^6 x \; dx = \frac{3.1.5.3.1}{10.8.6.4.2} \cdot \frac{\pi}{2} = \frac{3\pi}{512}$ ...(5)

From (3), (4) and (5), we have

$$\int_0^{2\pi} \sin^4 x \cos^6 x \; dx = 2.2. \frac{3\pi}{512} = \frac{3\pi}{128}.$$

Theorem 6·5. *If $f(x)$ has a primitive, then*

$$\int_{-a}^{a} f(x) \, dx = \begin{cases} 2\int_0^a f(x) \, dx, \text{ if } f(x) \text{ is an even function,} \\ 0, \text{ if } f(x) \text{ is an odd function.} \end{cases}$$

Proof. $\int_{-a}^{a} f(x) \, dx = \int_{-a}^{0} f(x) \, dx + \int_0^a f(x) \, dx$ ...(1)

Put $-x=t$ (so that $dx=-dt$) in the first integral on the right, Also $t=a$ when $x=-a$, and $t=0$, when $x=0$.

$$\therefore \int_{-a}^{0} f(x) dx = -\int_a^0 f(-t) \, dt = \int_0^a f(-t) \, dt = \int_0^a f(-x) \, dx \quad ...(2)$$

From (1) and (2) we have

$$\int_{-a}^{a} f(x) \, dx = \int_0^a f(-x) \, dx + \int_0^a f(x) \, dx \quad ...(3)$$

If $f(x)$ is an even function, $f(-x)=f(x)$ and (3) reduces to

$$\int_{-a}^{a} f(x) \, dx = \int_0^a f(x) \, dx + \int_0^a f(x) \, dx = 2\int_0^a f(x) \, dx \quad ...(4)$$

If $f(x)$ is an odd function, $f(-x)=-f(x)$ and (3) reduces to

$$\int_{-a}^{a} f(x) \, dx = \int_0^a -f(x) \, dx + \int_0^a f(x) \, dx = 0. \quad ...(5)$$

Example 15. *Evaluate* :

$$\int_0^{\pi/2} \frac{\cos x}{\cos x+\sin x}\,dx.$$

Solution.

Let $\quad f(x)=\dfrac{\cos x}{\cos x+\sin x}.$

$$\therefore \quad f(\pi/2-x)=\frac{\cos(\pi/2-x)}{\cos(\pi/2-x)+\sin(\pi/2-x)}$$

$$=\frac{\sin x}{\sin x+\cos x}$$

Since $\displaystyle\int_0^{\pi/2} f(x)\,dx=\int_0^{\pi/2} f(\pi/2-x)\,dx,$

therefore

$$I=\int_0^{\pi/2}\frac{\cos x}{\cos x+\sin x}dx=\int_0^{\pi/2}\frac{\sin x}{\sin x+\cos x}dx \qquad ...(1)$$

$$\therefore\ 2I=\int_0^{\pi/2}\frac{\cos x+\sin x}{\cos x+\sin x}dx=\int_0^{\pi/2}1.\,dx=\pi/2.$$

$\therefore \quad I=\pi/4.$

Remark. Note that we have added the two expressions for I as given in (1) to abtain a simple expression for I which could be easily evaluated.

Example 16. *Show that*

$$\int_0^{\pi/4} ln\,(1+\tan x)\,dx=\frac{\pi}{8}\,ln\,2.$$

Solution.

Let $\quad f(x)=ln\,(1+\tan x).$

$$f(\pi/4-x)=ln\,[1+\tan(\pi/4-x)],$$

$$=ln\left[1+\frac{1-\tan x}{1+\tan x}\right],$$

$$=ln\,\frac{2}{1+\tan x},$$

$$=\ln 2-\ln(1+\tan x),$$

$$=\ln 2-f(x). \qquad ...(1)$$

Integrating (1) over $[0,\pi/4]$, we have

$$\int_0^{\pi/4} f(\pi/4-x)\,dx=\int_0^{\pi/4} \ln 2\,dx-\int_0^{\pi/4} f(x)\,dx \qquad ...(2)$$

Since $\int_0^{\pi/4} f(\pi/4-x)\,dx=\int_0^{\pi/4} f(x)\,dx=\mathrm{I}$ (say),

therefore we may write (2) as

$$\mathrm{I}=\int_0^{\pi/4} \ln 2\,dx-\mathrm{I},$$

or

$$2\mathrm{I}=\int_0^{\pi/4} \ln 2\,dx=\frac{\pi}{4}\ln 2,$$

or

$$\mathrm{I}=\frac{\pi}{8}\ln 2.$$

Example 17. *Evaluate :*

$$\int_0^{\pi/2} \frac{x\sin x}{1+\cos^2 x}\,dx$$

Solution. Let

$$f(x)=\frac{x\sin x}{1+\cos^2 x},$$

$$f(\pi-x)=\frac{(\pi-x)\sin(\pi-x)}{1+\cos^2(\pi-x)},$$

$$=\frac{\pi\sin x}{1+\cos^2 x}-\frac{x\sin x}{1+\cos^2 x},$$

$$=\frac{\pi\sin x}{1+\cos^2 x}-f(x). \qquad ...(1)$$

Since $\int_0^{\pi} f(x)\,dx=\int_0^{\pi} f(\pi-x)\,dx+\mathrm{I}$ (say),

$\therefore$ Integrating both sides of (1), we have

$$I=\int_0^{\pi}\frac{\pi \sin x}{1+\cos^2 x}\,dx-I,$$

or $$2I=\pi\int_0^{\pi}\frac{\sin x}{1+\cos^2 x}\,dx \qquad ...(2)$$

Let $$g(x)=\frac{\sin x}{1+\cos^2 x}.$$

Then $$g(\pi-x)=\frac{\sin(\pi-x)}{1+\cos^2(\pi-x)}=\frac{\sin x}{1+\cos^2 x}=g(x),$$

$$\therefore \quad \int_0^{\pi} g(x)\,dx=2\int_0^{\pi/2} g(x)\,dx. \qquad ...(3)$$

From (2) and (3) we have

$$2I=2\pi\int_0^{\pi/2}\frac{\sin x\,dx}{1+\cos^2 x}. \qquad ...(4)$$

Putting $\cos x=t$, $-\sin x\,dx=dt$,

we have from (4),

$$I=\int_0^1\frac{dt}{1+t^2}=\pi\Big[\tan^{-1} t\Big]_0^1,$$

$$=\pi[\tan^{-1}1-\tan^{-1}0]$$
$$=\pi(\pi/4-0),$$
$$=\pi^2/4.$$

Example 18. *Show that*

$$\int_0^{\pi/2} \ln \sin x\,dx=-\frac{\pi}{2}\ln 2.$$

Solution. Let

$$f(x)=\ln \sin x. \quad \therefore \quad f\left(\frac{\pi}{2}-x\right)=\ln \cos x.$$

Since $$\int_0^{\pi/2} f(x)\,dx=\int_0^{\pi/2} f\left(\frac{\pi}{2}-x\right)dx,$$

$$\therefore \qquad \int_0^{\pi/2} \ln \sin x \; dx = \int_0^{\pi/2} \ln \cos x \; dx = \text{I (say)}$$

$$\therefore \qquad 2\text{I} = \int_0^{\pi/2} [\ln \sin x + \ln \cos x] dx,$$

$$= \int_0^{\pi/2} \ln (\sin x \cos x) \, dx,$$

$$= \int_0^{\pi/2} \ln \left(\frac{1}{2} \sin 2x \right) dx,$$

$$= \int_0^{\pi/2} \left(\ln \frac{1}{2} + \ln \sin 2x \right) dx,$$

$$= \int_0^{\pi/2} \left(\ln \frac{1}{2} \right) dx + \int_0^{\pi/2} \ln \sin 2x \, dx,$$

$$= \frac{\pi}{2} \ln \frac{1}{2} + \text{J (say)}, \qquad \ldots(1)$$

where $$\text{J} = \int_0^{\pi/2} \ln \sin 2x \, dx,$$

Putting $2x = t$, $2dx = dt$, we have

$$\text{J} = \frac{1}{2} \int_0^{\pi} \ln \sin t \, dt \qquad \ldots(2)$$

Now $f(t) = \ln \sin t = \ln \sin (\pi - t) = f(\pi - t)$.

$$\therefore \qquad \int_0^{\pi} f(t) \, dt = 2 \int_0^{\pi/2} f(t) \, dt = 2\text{I},$$

or $$\text{J} = \frac{1}{2} \int_0^{\pi} f(t) dt = \frac{1}{2} \cdot 2\text{I} = \text{I}. \qquad \ldots(3)$$

From (1) and (3), we have

$$2I=\frac{\pi}{2}\ln\frac{1}{2}+I,$$

or

$$I=\frac{\pi}{2}\ln\frac{1}{2}=-\frac{\pi}{2}\ln 2.$$

EXERCISES 6 (*e*)

Show that

1. $\displaystyle\int_0^{\pi/2}\frac{\sqrt{\cos x}}{\sqrt{\cos x}+\sqrt{\sin x}}\,dx=\frac{\pi}{4}.$

2. $\displaystyle\int_0^{\pi/2}\ln\tan x\,dx=0.$

3. $\displaystyle\int_0^{\pi/2}\frac{\cos x-\sin x}{1+\cos x\sin x}\,dx=0.$

4. $\displaystyle\int_0^{\pi}x\sin^3 x\,dx=\frac{2\pi}{3}.$

5. $\displaystyle\int_0^{\pi}x\sin^4 x\,dx=\frac{3}{16}\pi^2.$

6. $\displaystyle\int_0^{\pi}x\cos^2 x\sin^4 x\,dx=\frac{\pi^2}{32}.$

7. $\displaystyle\int_0^{\pi}x\sin^4 x\cos^6 x\,dx=\frac{3\pi^2}{512}.$

8. $\displaystyle\int_0^{\pi}\ln(1-\cos x)\,dx=\pi\ln\left(\frac{1}{2}\right).$

9. $\displaystyle\int_0^{\pi/2}\ln(\cot x+\tan x)\,dx=\pi\ln 2.$

10. $$\int_0^{\pi} x \ln \sin x \, dx = \frac{1}{2}\pi^2 \ln\left(\frac{1}{2}\right).$$

11. $$\int_0^{\pi/2} \frac{\cos^2 x}{1+\cos x \sin x}\, dx = \frac{\pi}{3\sqrt{3}}.$$

12. $$\int_0^{\pi} \frac{x \tan x}{\sec x+\tan x}\, dx = \frac{\pi(\pi-2)}{2}.$$

13. $$\int_0^{\pi} \frac{x\, dx}{4\cos^2 x+9\sin^2 x} = \frac{\pi^2}{12}.$$

14. $$\int_0^{1} \frac{\ln(1+x)}{1+x^2}\, dx = \frac{\pi}{8} \ln 2.$$

15. $$\int_0^{\pi} \frac{x}{1+\sin x}\, dx = \pi.$$

6·8. APPLICATION TO PLANE AREAS

We have seen how as an application of the fundamental theorem of the integral calculus, we can evaluate definite integrals of various functions. Since the areas of certain plane regions can be expressed as definite integrals, we are now in a position to determine these areas with the techniques available to us. This provides justification for defining the intuitive concept of an area as definite integral.

You will find that sometimes it is not only convenient but also necessary to sketch the curve before determining its area.

Example 19. *Find the area of the region bounded by the straight line $x-2y+2=0$, the ordinates $x=1$, $x=2$, and the x-axis.*

Solution. The equation of the straight line is

$$x-2y+2=0 \quad \text{or} \quad y=\tfrac{1}{2}(x+2).$$

Thus the region whose area is required, is bounded by the curve $y=f(x)$ where $f(x)=\frac{1}{2}(x+2)$, the ordinates $x=1$, $x=2$, and the x-axis.

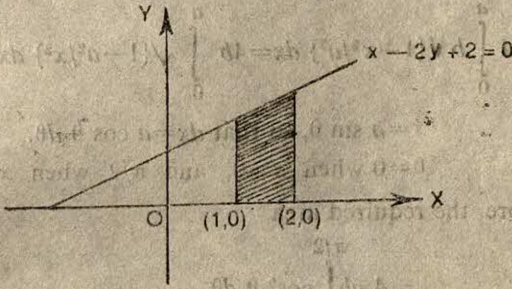


Fig. 6·7.

$\therefore$ The required area is $\int_1^2 \frac{1}{2}(x+2)\,dx.$

Now $\int_1^2 \frac{1}{2}(x+2)\,dx = \frac{1}{2}\int_1^2 (x+2)\,dx.$

$$=\frac{1}{2}\left[\frac{x^2}{2}+2x\right]_1^2$$

$$=\frac{1}{2}\left[\frac{1}{2}\cdot 4+2\cdot 2-\frac{1}{2}-2\right].$$

$$=1\tfrac{3}{4}.$$

Example 20. *Find the area of the ellipse*

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1.$$

Solution. The ellipse is symmetrical about the axes. Therefore, the area enclosed within the ellipse is 4 times the area OAB of the ellipse in the first quadrant. This area is enlosed by the ellipse, the ordinates $x=0$, $x=a$, and the x-axis. The equation of the ellipse is given by

$$y^2=b^2(1-x^2/a^2),$$

or $\quad y=f(x)=b\sqrt{\{1-x^2/a^2\}}.$

(y is positive for points on the ellipse in the first quadrant.)

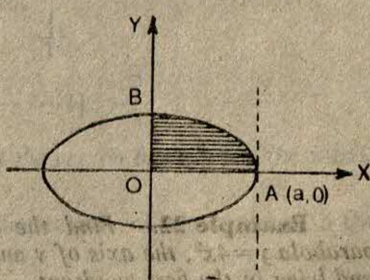


Fig. 6·8.

Therefore, the required area is

$$4\int_0^a b\,\sqrt{\{1-x^2/a^2\}}\,dx=4b\int_0^a \sqrt{\{1-a^2/x^2\}}\,dx.$$

Put $\qquad x=a\sin\theta$, so that $dx=a\cos\theta\,d\theta$.

Also, $\qquad \theta=0$ when $x=0$, and $\pi/2$ when $x=a$.

Therefore, the required area

$$=4\,ab\int_0^{\pi/2}\cos^2\theta\,d\theta,$$

$$=\frac{4ab}{2}\int_0^{\pi/2}(1+\cos 2\theta)\,d\theta,$$

$$=2ab\left[\theta+\frac{\sin 2\theta}{2}\right]_0^{\pi/2}=2ab\,.\,\frac{\pi}{2}=\pi ab\cdot$$

Example 21. *Find the area between the curve $y=x^2$, the x-axis, and the lines $x=0$ and $x=6$.*

Solution.

The required area is given by

$$\int_0^6 x^2\,dx=\left[\frac{x^3}{3}\right]_0^6=72.$$

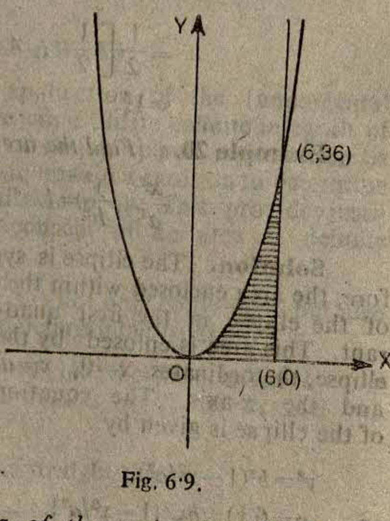


Fig. 6·9.

Example 22. *Find the area of the region bounded by the parabola $y=4x^2$, the axis of y and the two abscissae $y=1$ and $y=4$, and lying in the first quadrant.*

Solution.

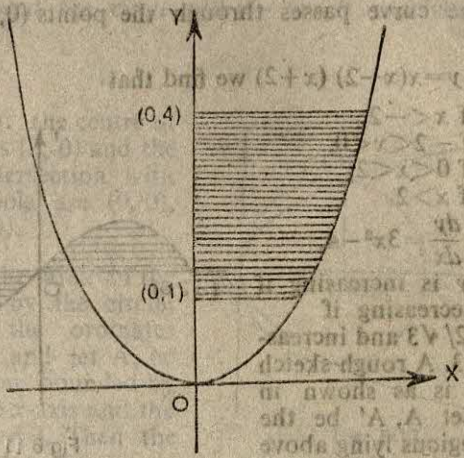


Fig. 6·10.

It may be noted that the area of the region bounded by the curve $x=f(y)$, the abscissae $y=c$, $y=d$, and the y-axis is given by

$$\int_c^d f(y)\, dy.$$

In the given case $x=f(y)=\frac{1}{2}\sqrt{y}$, taking the positive sign because in the first quadrant, x is positive.

Therefore, the required area is

$$\int_1^4 \frac{1}{2}\sqrt{y}\, dy=\frac{1}{2}\int_1^4 \sqrt{y}\, dy,$$

$$=\frac{1}{2}\left[\frac{2}{3}y^{3/2}\right]_1^4$$

$$=\frac{1}{2}\cdot\frac{2}{3}[8-1]=\frac{7}{3}.$$

Example 23. *Find the area of the region bounded by the* **x-axis** *and the curve $y=x^3-4x$.*

Solution. The points of intersection of the curve with the x-axis are given by

$$x^3-4x=0,$$

i.e., $x=0, 2, -2.$

Thus, the curve passes through the points (0, 0), (2, 0) and (−2, 0).

Writing $y=x(x-2)(x+2)$ we find that

$y<0$ if $x<-2$,
$y>0$ if $-2<x<0$,
$y<0$ if $0<x<2$,
$y>0$ if $x>2$.

Since $\frac{dy}{dx}=3x^2-4$,

we find that y is increasing if $x<-2/\sqrt{3}$, decreasing if $-2/\sqrt{3}<x<2/\sqrt{3}$ and increasing if $x>2/\sqrt{3}$. A rough sketch of the curve is as shown in Fig. 6·11. Let A, A′ be the areas of the regions lying above and below the x-axis respectively.

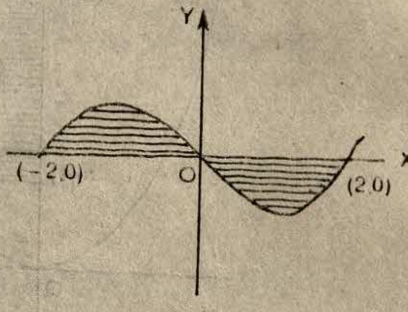


Fig 6 11

Then $A=\int_{-2}^{0}(x^3-4x)\,dx=\left[\frac{x^4}{4}-2x^2\right]_{-2}^{0}=4.$

Notice that $f(x)=x^3-4x\geqslant 0$ throughout the interval $[-2, 0]$. Also, $f(x)\leqslant 0$ throughout $[0, 2]$.

$$\therefore \quad A'=\int_0^2 -f(x)\,dx=-\int_0^2 (x^3-4x)\,dx.$$

$$=-\left[\frac{x^4}{4}-2x^2\right]_0^2=4.$$

Thus the required area $=A+A'=4+4=8.$

Remark.

While evaluating $\int_a^b f(x)\,dx$, it must be ensured that $f(x)\geqslant 0$ throughout $[a, b]$.

If $f(x)\leqslant 0$ over $[a, b]$, the area would be given by

$\int_a^b -f(x)\,dx$ as explained earlier.

Example 24. *Find the area of the region included between the parabola $y^2=3x$ and the circle $x^2+y^2-6x=0$ and lying in the first quadrant.*

Solution.

Notice that the centre of the given circle is (3, 0) and the points of its intersection with the given parabola are (0, 0), (3, 3) and (3, −3).

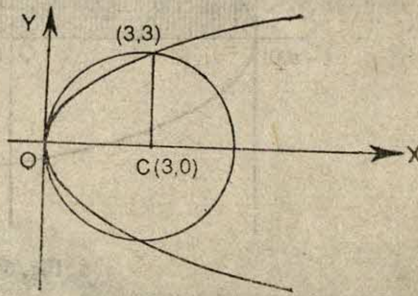


Fig. 6·12.

Let A_1 be the area of the region bounded by the circle, the x-axis and the ordinates $x=0$ and $x=3$, and let A_2 be the area of region bounded by the parabola, the x-axis and the ordinates $x=0$, $x=3$. Then the required area is A_1-A_2.

Now, $$A_1=\int_0^3 \sqrt{(6x-x^2)}\,dx,$$

$$=\int_0^3 \sqrt{\{3^2-(x-3)^2\}}\,dx,$$

$$=9\pi/4$$ (as can be seen by making the substitution $x-3=3\sin\theta$.)

Also, $$A_2=\int_0^3 \sqrt{3x}\,dx=6.$$

Therefore the required area

$$=\frac{9\pi}{4}-6=\frac{3}{4}(3\pi-8).$$

Example 25. *Find the area of the region included between the parabolas $y^2=4a(x+a)$ and $y^2=4b(b-x)$.*

Solution. The parabolas intersect at the points where

$$4a(x+a)=4b(b-x),$$

or where $$x=b-a.$$

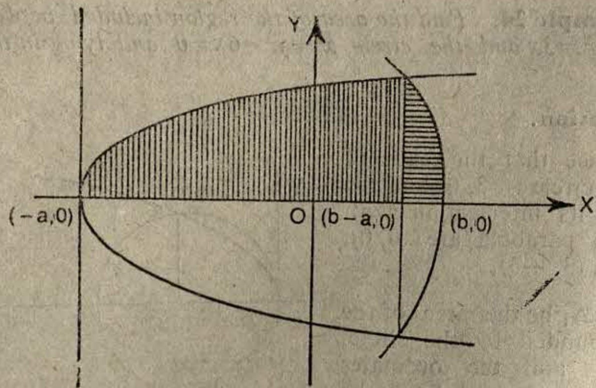


Fig. 6·13.

Obviously, because of the symmetry of the parabolas about the x-axis, the required area is twice the sum of the areas A_1 and A_2, where A_1 is the area of the region bounded by the parabola $y^2=4a(x+a)$, the x-axis and the ordinates $x=-a$ and $x=b-a$ and A_2 is the area of the region bounded by the parabola $y^2=4b(b-x)$, the x-axis and the ordinates $x=b-a$, $x=b$.

Therefore the required area is

$$=2\left[\int_{-a}^{b-a}\sqrt{\{4a(x+a)\}}\,dx+\int_{b-a}^{b}\sqrt{\{4b(b-x)\}}\,dx,\right.$$

$$=2\left[2\sqrt{a}\cdot\frac{2}{3}\Big[(x+a)^{3/2}\Big]_{-a}^{b-a}\right.$$

$$\left.-2\sqrt{b}\cdot\frac{2}{3}\Big[(b-x)^{3/2}\Big]_{b-a}^{b}\right],$$

$$=\frac{8}{3}(a+b)\sqrt{ab}.$$

EXERCISE 6 (*f*)

1. Find the area bounded by the straight line $x+y=10$, the ordinates $x=1$, $x=8$, and the x-axis.

2. Find the area bounded by the curve $y=e^x$, the ordinates $x=0$, $x=3$, and the x-axis.

3. Find the area bounded by the curve $y=x^3$, the ordinates $x=0$, $x=2$ and the x-axis.

4. Find the area bounded by the curve $y=\ln x$, the ordinates $x=1$, $x=2$, and the x-axis.

5. Find the area of the circle $x^2+y^2=a^2$.

6. Find the area bounded by the parabola $y^2=4ax$ and its latus rectum.

7. Find the area bounded by the curve $y=\sin x$, the ordinates $x=0$, $x=2$, and the x-axis.

8. Find the area bounded by the curve

$y=x(x-3)(x-5)$,

the ordinates $x=0$, $x=5$, and the x axis.

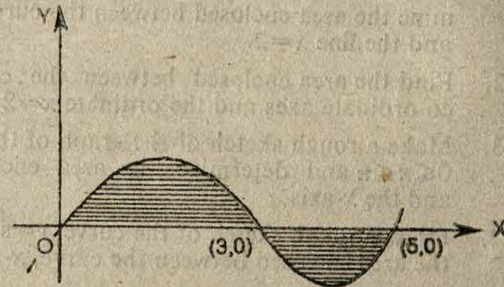


Fig. 6·14.

9. Find the area bounded by the parabola $y=x^2$ and the straight line $y=2x$.

10. Find the area included between the circle $x^2+y^2=2ax$ and the parabola $y^2=ax$, lying above the x-axis.

11. Find the area included between the parabolas $y^2=4ax$ and $x^2=4ay$. *(Roorkee Entrance, 1980)*

12. Find the area in the first quadrant bounded by the parabola $y^2=4x$, the circle $x^2+y^2=5$, and the x-axis.

13. Find the area included between the parabolas $y^2=-4(x-1)$ and $y^2=-2(x-2)$.

14. Find the area cut off from the curve $y=x(2-x)$ by the line $2y=x$.

15. Draw a rough sketch of the curve $\frac{x^2}{4}+\frac{y^2}{9}=1$ and evaluate the area of the region under the curve and above the x-axis. *(A.I S.S.C.E., 1987)*

16. Draw a rough graph of the curve $y=4-x^2$ and find the area enclosed by the curve and the lines $x=0$, $x=2$ and the x-axis. *(D.B.S S C.E , 1989)*

17. Draw a graph of $y^2+1=x$, $x\leqslant 2$ and find the area enclosed by the curve and the ordinate $x=2$. *(D B.S S.C E , 1986)*

18. Sketch the rough graph of $y=4\sqrt{x-1}$, $1\leqslant x\leqslant 3$ and evaluate the area between the curve, x-axis and the line $x=3$. *(A.I.S.S.C.E., 1985)*

19. Draw the rough sketch of the curve $y=\sqrt{3x+4}$ and find the area under the curve, above the x-axis between $x=0$ and $x=4$. [*A.I.S.SC.E., 1984*)

20. Draw a rough sketch of the curve $y=\sqrt{x-1}$ in the interval $[1, 5]$ and find the area under the curve, above the x-axis and between the lines $x=1$ and $x=5$.

21. Make a rough sketch of the graph of $y=\sqrt{6x+4}$ and determine the area enclosed between the curve, the co-ordinate axes and the line $x=2$. (*D.B.S.S.C.E., 1988*)

22. Find the area enclosed between the curve $y^2=x^2(4-x^2)$, the co-ordinate axes and the ordinate $x=2$. (*A.I.S.S.C.E., 1986*)

23. Make a rough sketch of the graph of the function $y=3 \sin x$, $0 \leqslant x \leqslant \pi$ and determine the area enclosed between the curve and the x-axis. (*D.B.S.S.C.E., 1984*)

24. Draw a rough sketch of the curve $y=\sin^2 x$, $x \in [0, \pi/2]$. Find the area enclosed between the curve, x-axis and the line $x=\pi/2$. (*D.B.S S.C.E., 1987*)

25. Make a rough sketch of the graph of the function $y=\dfrac{4}{x^2}$, $1 \leqslant x \leqslant 3$ and find the area enclosed between the curve, the x-axis, and the lines $x=1$ and $x=3$. (*A.I.S.S.C.E. 1984*)

26. Find the area bounded by the curve $y=(x-1)(x-2)(x-3)$ lying between the ordinates $x=0$ and $x=3$. (*Roorkee, 1985*)

27. Find the area in the xy-plane enclosed by the curve $a^2y^2=x^2(a^2-x^2)$.

TEST YOUR UNDERSTANDING VI

In each of the following problems four alternatives are given out of which only one is correct. Put a tick mark ($\surd$) against the correct alternative :

1. $\int_0^1 x^2\, dx$ equals

(*a*) 1 (*b*) 2

(*c*) $\frac{1}{2}$ (*d*) $\frac{1}{3}$.

2. $\int_0^2 e^{2x}\, dx$ equals

(*a*) e^4 (*b*) $\frac{1}{2} e^4$

(*c*) $\frac{1}{2}(e^4-1)$ (*d*) $2\, e^4$.

3. $\int_0^1 \frac{dx}{1+x}$ equals

(a) $ln\ 2$ (b) $\frac{1}{2}$

(c) $ln\ \frac{1}{2}$ (d) 2.

4. $\int_0^1 \frac{dx}{1+x^2}$ equals

(a) $\frac{1}{2}$ (b) $ln\ 2$

(c) $\pi/4$ (d) $\pi/2$.

5. $\int_0^1 \frac{dx}{\sqrt{(4-x^2)}}$ equals

(a) $\frac{\pi}{2}$ (b) $\frac{1}{2}\ ln\ \frac{3}{2}$.

(c) $\frac{\pi}{4}$ (d) $\frac{\pi}{6}$.

6. $\int_2^3 \frac{dx}{x^2-1}$ equals

(a) $ln\ 2$ (b) $\frac{1}{2}\ ln\ \frac{3}{2}$

(c) $ln\ 3$ (d) $2\ ln\ 6$.

7. $\int_0^{\pi} \sin^4 x\, dx$ equals

(a) $\frac{\pi}{16}$ (b) $\frac{3\pi}{16}$

(c) $\frac{3\pi}{8}$ (d) π.

8. $\int_0^{2\pi} \cos^3 x\, dx$ equals

(a) 0 (b) 1

(c) $\frac{\pi}{2}$ (d) $\frac{2}{3}$.

9. The area bounded by $y=x^3-3x$, the x-axis and the ordinates $x=-1$, $x=1$ equals

(a) $\frac{5}{4}$ (b) $\frac{5}{2}$

(c) 5 (d) 0.

10. The area bounded by $y=\sin x$, the x-axis, between $x=-\pi/2$ and $x=\pi/2$ is

(a) 0 (b) 1

(c) 2 (d) π.

REVIEW EXERCISE VI

Evaluate the following integrals :

1. $\int_0^{\sqrt{2}} \sqrt{2-x^2}\, dx.$ *(A.I.S.S.C.E., 1984)*

2. $\int_{-1}^{1} ln \left(\frac{2-x}{2+x}\right) dx.$ *(Roorkee Entrance, 1986)*

3. $\int_0^{\pi/2} \cos 2x \cos 3x\, dx.$ *(D.B.S.S.C.E., 1987)*

4. $\int_0^{\pi/2} \frac{\cos x\, dx}{1+\sin^2 x}.$ *(D.B.S.S.C.E., 1984)*

5. $\int_0^{\pi/6} (2+3x^2) \cos 3x\, dx.$ *(D.B.S.S.C.E., 1985)*

6. $\int_{-a}^{a} \sqrt{\frac{a-x}{a+x}}\, dx.$ *(A.I.S.S.C.E., 1986)*

7. $\int_0^{\pi/2} \frac{\sin x \cos x}{1+\sin^4 x}\, dx$ *(A.I.S.S.C.E., 1988)*

8. $\int_0^{1} \cos^{-1} x\, dx$ *(D.B.S.S.C.E., 1988)*

9. $\int_0^{\pi/4} \sqrt{1+\sin 2x}\, dx$ (*A.I.S.S.C.E., 1985*)

10. $\int_0^{\pi/2} \frac{\cos x\, dx}{1+\cos x+\sin x}$ (*Roorkee, 1989*)

11. Evaluate $\int_1^2 e^{3x}\, dx$ by expressing it as the limit of a sum.

12. Find the limit, when n tends to infinity, of the series :

(*a*) $\sum_{r=1}^{n} \frac{n^3}{(n^2+r^2)(n^2+2r^2)}$

(*b*) $\sum_{r=1}^{n} \frac{\sqrt{n}}{(9n+40r)^{3/2}}$

13. Show that

$$\int_0^{\pi} \frac{x}{a \cos^2 x+b^2 \sin^2 x}\, dx=\frac{\pi^2}{2ab}$$

14. Show that

(*a*) $\int_a^b f(a+b-x)\, dx=\int_a^b f(x)\, dx$

(*b*) $\int_0^{b-c} f(x+c)\, dx=\int_c^b f(x)\, dx$

15. Show that

(*a*) $\int_{-a}^{a} f(x^2)\, dx=2\int_0^a f(x^2)\, dx$

(*b*) $\int_{-a}^{a} xf(x^2)\, dx=0.$

16. Show that

(a) $$\int_0^{\pi} xf(\sin x)\,dx=\frac{\pi}{2}\int_0^{\pi} f(\sin x)\,dx$$

(b) $$\int_a^{\pi-a} x f(\sin x)\,dx=\frac{\pi}{2}\int_a^{\pi-a} f(\sin x)\,dx.$$

17. If f is a periodic function with period p (*i.e.*, $f(x+p)=f(x)$ for all x), prove that

$$\int_0^{np} f(x)\,dx=n\int_0^{p} f(x)\,dx,$$

where n is any integer.

18. Show that if n be any positive integer,

$$\int_0^{n\pi} f(\cos^2 x)\,dx=n\int_0^{\pi} f(\cos^2 x)\,dx.$$

19. Show that

$$\int_{-a}^{a} f(x)f(-x)\,dx=2\int_0^{a} f(x)f(-x)\,dx.$$

20. Find the area between the curve

$x^2/a^2+y^2/b^2=1$ and the x-axis between $x=0$ and $x=a$. Draw a rough graph of the curve also. (*D.B.S.S.C.E., 1985*)

21. Find the area of a loop between the curve $y=a\sin x$ and the x-axis. (*Roorkee Entrance, 1989*)

22. Find the area bounded by the curve $y=x\sin x$ and the x-axis between $x=0$ and $x=2\pi$. (*Roorkee Entrance, 1981*)

23. Find the area of the region bounded by the curves $y=\log_e x$, $y=\sin^4 \pi x$, and $x=0$. (*Roorkee Entrance, 1987*)

24. Find the area included between the parabola $y=\dfrac{x^2}{4a}$ and the witch $y=\dfrac{8a^3}{x^2+4a^2}$. (*Roorkee Entrance, 1983*)

25. Find the area under the curve $x^4=a^2(x^2-y^2)$ above the x-axis and between the lines $x=0$ and $x=a$. (*A.I.S.S.C.E., 1986*)

26. Draw a rough sketch of the curve

$$y=\frac{x}{\pi}+2\sin^2 x$$

and find the area between the x-axis, the curve and the ordinates $x=0$ and $x=\pi$. *(A.I.S.S.C.E., 1989)*

27. Find the area between the curve $y=2x^4-x^2$, the x-axis and the ordinates of the two minima of the curve. *(Roorkee Entrance, 1988)*

SUMMARY

1. $$\int_a^b f(x)dx=\lim_{\substack{n\to\infty\\ h\to 0\\ nh=b-a}} [h\{f(a)+f(a+h)+\ldots+f(a+\overline{n-1}h)\}]$$

$$=\lim_{\substack{n\to\infty\\ h\to 0\\ nh=b-a}} [h\{f(a+h)+f(a+2h)+\ldots+f(a+nh)\}]$$

2. **The fundamental theorem of integral calculus :**

Let $f(x)$ be defined and continuous over the closed interval $[a, b]$. If $F(x)$ is a primitive of $f(x)$, then

$$\int_a^b f(x)\,dx=F(b)-F(a).$$

3. $$\int_a^a f(x)\,dx=0.$$

4. $$\int_b^a f(x)\,dx=-\int_a^b f(x)\,dx.$$

5. $$\int_a^b [f(x)+g(x)]\,dx=\int_a^b f(x)\,dx+\int_a^b g(x)\,dx.$$

6. $\int_a^b k\,f(x)\,dx=k\int_a^b f(x)\,dx$, where k is a constant.

7. $$\int_a^b f(x)\,dx=\int_a^c f(x)\,dx+\int_c^b f(x)\,dx. \quad (a<c<b)$$

8. $\displaystyle\int_0^a f(x)\,dx = \int_0^a f(a-x)\,dx.$

9. $\displaystyle\int_0^{2a} f(x)\,dx = \int_0^a f(x)\,dx + \int_0^a f(2a-x)\,dx.$

10. $\displaystyle\int_0^{2a} f(x)\,dx = \begin{cases} 2\displaystyle\int_0^a f(x)\,dx, & \text{if } f(2a-x)=f(x) \\ 0, & \text{if } f(2a-x)=-f(x) \end{cases}$

11. $\displaystyle\int_{-a}^{a} f(x)\,dx = \begin{cases} 2\displaystyle\int_0^a f(x)\,dx, & \text{if } f \text{ is an even function,} \\ 0, & \text{if } f \text{ is an odd function.} \end{cases}$

12. **Substitution Rule.** Let $f(x)$ be a function which has a primitive. Let $x=\phi(t)$ be a function and α, β be real numbers such that $\phi(\alpha)=a$, $\phi(\beta)=b$. If $\phi(t)$ possesses a continuous derivative on $[\alpha, \beta]$, then

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f(\phi(t))\,\phi'(t)\,dt$$

13. $\displaystyle\lim_{\substack{n\to\infty \\ h\to 0 \\ nh=1}} h\{f(0)+f(h)+f(2h)+\ldots+f(\overline{n-1}\;h)\} = \int_0^1 f(x)\,dx.$

14. $\displaystyle\lim_{\substack{n\to\infty \\ h\to 0 \\ nh=1}} \{h(f(h)+f(2h)+\ldots+f(nh)=\}\int_0^1 f(x)\,dx.$

15. $\displaystyle\int_0^{\pi/2} \sin^m x \cos^n x\,dx$

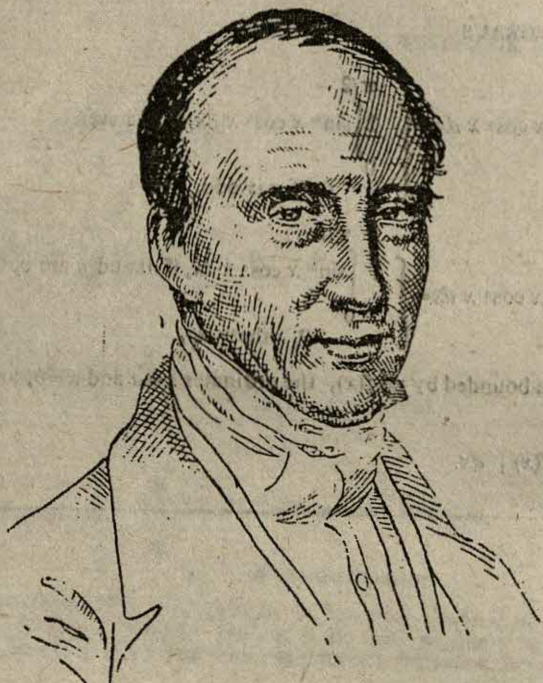
$$= \frac{(m-1)(m-3)\ldots(n-1)(n-3)\ldots}{(m+n)(m+n-2)\ldots} \times k,$$

the three sets of factors starting with $m-1$, $n-1$ and $m+n$, and diminishing by 2 at a time, end up with either 1 or 2 according as the first factor of the set is odd or even, $k=\pi/2$ if m and n are both even, and $k=1$ if at least one of m and n is odd.

16. $$\int_0^{\pi} \sin^m x \cos^n x \, dx = \begin{cases} 2\displaystyle\int_0^{\pi/2} \sin^m x \cos^n x \, dx, & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

17. $$\int^{2\pi} \sin^m x \cos^n x \, dx = \begin{cases} 4\displaystyle\int_0^{\pi/2} \sin^m x \cos^n x \, dx, & \text{if } m \text{ and } n \text{ are both even.} \\ 0, & \text{otherwise.} \end{cases}$$

18. The area bounded by $y=f(x)$, the ordinates $x=a$ and $x=b$, and the x-axis is $\displaystyle\int_a^b |f(x)| \, dx$.



AUGUSTIN-LOUIS CAUCHY (1789-1957)

Augustin-Louis Cauchy was born in Paris on August 21, 1789. He was educated at the Ecole Polytechnique, where he was a student of Lagrange and Laplace.

In 1811, Cauchy submitted his first Memoir on the theory of polyhedra. The history of determinants begins in 1812 with Cauchy. It was in this connection that he proved the result on parity of permutations that you would read soon. He began the study of the theory of substitutions in the middle 1840's, which later developed into the theory of finite groups.

Modern mathematics is indebted to Cauchy for the introduction of rigour into Mathematical Analysis. In 1821, Cauchy published his course of lectures on analysis which he gave at the ecole Polytechnique. This is the work which set standards in rigorous mathematics. Our present day definitions of limit and continuity, derivative as the limit of a difference quotient, the definite integeral as the limit of a sum, are substatially the same as given by Cauchy in this course of lectures.

Cauchy's mathematical productivity was incredible. He started two journals of his own for the publication of his expository and original work in pure and applied mathematics. These works were eagerly bought and studied, and did much to reform mathematical taste before 1860. During the last 19 years of his life, Cauchy produced over 500 papers on all branches of the mathematics including mechnics, physics and astronomy. Many of these works were long treatises. In fact, Cauchy invited a lot of criticism for over-production and hasty composition. His total output in 789 papers filling twenty-four large volumes. However, his reputation as mathematician has risen steadily.

He died rather unexpectedly in his sixty eighth year on May 23, 1857.

CHAPTER 7

Differential Equations

7·1. INTRODUCTION

Differential equations are of utmost importance in physics and engineering. Many physical laws when expressed mathematically take the form of differential equations. In chemistry, biology, economics and various other disciplines also we come across situations where mathematical formulation gives rise to differential equations. In the present chapter we shall study some simple differential equations and their applications.

An equation in which at least one term contains $\frac{dy}{dx}$, $\frac{d^2y}{dx^2}$ etc. is called a differential equation. For example,

$$\frac{dy}{dx}+4y=3x, \qquad \text{...(1)}$$

$$\left(\frac{dy}{dx}\right)^2+4x^2\frac{dy}{dx}-y=3x; \qquad \text{...(2)}$$

$$\frac{d^2y}{dx^2}+\frac{dy}{dx}-y=e^x, \qquad \text{...(3)}$$

$$\frac{d^2y}{dx^2}+\left(\frac{dy}{dx}\right)^3-3y=\sin x, \qquad \text{...(4)}$$

$$\left(\frac{d^2y}{dx^2}\right)^3+\frac{dy}{dx}-3y=\sin x, \qquad \text{...(5)}$$

are all differential equations.

7·1·1. Order and degree of a differential equation

In the illustrations given above, (1) and (2) contain only the first derivative but do not contain any higher order derivative. Such equations are called differential equations of the first order. Equations (3), (4) and (5) contain derivatives of the first and second order but do not contain any derivative of a higher order. Because of this we say that these differential equations are of the second order.

Definition 7·1. *A differential equation is said to be of order n, if the nth order derivative is the highest derivative of y in that equation.*

Let us once again consider the illustrations (1)—(5) above. Equations (1) and (2) are both of the first order. But they differ in one respect. Whereas the highest power of $\frac{dy}{dx}$ in equation (1) is one, the highest power of $\frac{dy}{dx}$ in equation (2) is *two*. We expres this difference by saying that equation (1) is of the first degree and equation (2) is of the second degree.

Definition. 7·2. *The degree of a differential equation is the power of the derivative of the highest order in the equation when the equation has been expressed in a form so that it is free of radicals and fractions so far as the derivatives are concerned.*

In the illustrations above, equations (3) and (4) are of the first degree, and equation (5) is of the third degree.

7·1·2. General and particular solution of a differential equation

A function

$$y=f(x)$$

is called a *solution* of a given first order differential equation on some interval I if it is defined and differentiable at every point of I and is such that the equation becomes an identity when y and $\frac{dy}{dx}$ are replaced by f and f' respectively.

For example, the function

$$y=f(x)=e^{3x}$$

is a solution of the first-order differential equation

$$\frac{dx}{dx}=3y$$

for all x, because

$$f'(x)=3e^{6x},$$

and by substituting $y=e^{3x}\frac{dy}{dx}=3e^{3x}$, the equation reduces to the identity

$$3e^{3x}=3e^{3x}.$$

Remark. Sometimes a solution of a given differential equation turns out to be an implicit function, *i.e.*, a function of the form $F(x, y)=0$ instead of an implicit function $y=f(x)$. For example, it can be easily verified that $x^2-y^2=1$ $(y\neq 0)$ is a solution of the differential equation

$$\frac{dy}{dx}=\frac{x}{y}.$$

Example 1, *Verify that each of the functions*

$y=\cos x$, $y=\cos x-1$, $y=\cos x+2$,

$y=\cos x+C$ *(where C is any real number whatever) is a solution of the differential equation*

$$\frac{dy}{dx}+\sin x=0.$$

Solution.

Let $f(x)=\cos x$, $g(x)=\cos x-1$, $h(x)=\cos x+2$.

$k(x)=\cos x+C$.

Then

$f'(x)=-\sin x$, so that $f'(x)+\sin x=0$,

$g'(x)=-\sin x$, so that $g'(x)+\sin x=0$,

$h'(x)=-\sin x$, so that $h'(x)+\sin x=0$,

$k'(x)=-\sin x$, so that $k'(x)+\sin x=0$.

showing that $y=f(x)$, $y=g(x)$, $y=h(x)$ and $y=k(x)$ are all solutions of the first-order differential equation

$$\frac{dy}{dx}+\sin x=0.$$

Example 2. *Verify that each of the functions*

$y=e^{2x}$, $y=3e^{2x}$, $x=-5e^{2x}$, $y=C\,e^{2x}$ (where C is any real number whatever).

is a solution of the differential equation

$$\frac{dy}{dx}=2\,y.$$

Solution :

(*i*) Let $y=e^{2x}$, so that $\dfrac{dy}{dx}=2e^{2x}=2y$.

(*ii*) Let $y=3e^{2x}$, so that $\dfrac{dy}{dx}=3(2e^{2x})=3y$.

(*iii*) Let $y=-5e^{2x}$, so that $\dfrac{dy}{dx}=-5(2e^{2x})=-5y$.

(*iv*) Let $y=Ce^{2x}$, so that $\dfrac{dy}{dx}=C(2e^{2x})=2y$.

Thus $y=e^{2x}$, $y=3e^{2x}$, $y=-5e^{2x}$, and $y=Ce^{2x}$ are solutions of the given differential equation.

From the above examples we find that a first-order differential equation may have more than one (even infinitely many !) solutions. All these solutions can be represented by a single formula involving an arbitrary constant, say C. Observe that all the solutions in

Example 1 can be represented by $y=\cos x+C$, where C is an arbitrary constant. Similarly, all the solutions of the differential equation given in Example 2 can be represented by $y=Ce^{2x}$ where c is an arbitrary constant.

A solution of a first-order differential equation containing an arbitrary constant is called the *general solution.*

A solution obtained from the general solution by giving a particular value of the arbitrary constant is called a *particular solution.*

In Example 1, $y=\cos x+C$ (where C is an arbitrary constant) is the general solution of the given differential equation, and $y=\cos x$, $y=\cos x-1$, $y=\cos x+2$ are all particular solutions obtained by giving to C the values 0, -1 and 2 respectively.

In example 2, $y=Ce^{2x}$ (where C is an arbitrary constant) is the general solution of the given differential equation and

$y=e^{2x}$, $y=3e^{2x}$, $y=-e^{2x}$ are all particular solutions.

7·2. FORMATION OF A DIFFERENTIAL EQUATION

Before trying to learn the technique of solving some simple first-order differential equations, let us see as to how differential equations arise very naturally in various situations.

Illustrations.

(*a*) **Family of all lines with a given slope**

If $y=f(x)$ be the equation of a straight line with slope m, then we must have

$$\frac{dy}{dx}=m. \qquad \text{...(1)}$$

(1) is the differential equation of all straight lines having slope m.

(*b*) **Family of all straight lines passing through the origin.**

Let $y=f(x)$ be the equation of a straight line passing through the origin. If (x, y) be any point on the line, its slope will be y/x. Therefore we must have

$$\frac{dy}{dx}=\frac{y}{x} \qquad \text{...(2)}$$

at every point of every such line.

Therefore (2) is the differential equation of all straight lines passing through the origin.

(*c*) **Family of all curves, such that for each curve the normal at each point passes through the origin.**

Let $y=f(x)$ be the equation of a curve satisfying the given

condition, and let (x, y) be any point on the curve. The slope of the normal will be y/x, so that we must have

$$\left(\frac{y}{x}\right)\left(\frac{dy}{dx}\right)=-1,$$

or

$$y\frac{dy}{dx}+x=0,$$

which is the desired differential equation.

(*d*) **Motion in a straight line with uniform acceleration.**

Suppose that a particle moves in a straight line with a uniform acceleration f. The rate of change of velocity v at time t is $\frac{dv}{dt}$, and therefore the motion of the particle satisfies the differential equation,

$$\frac{dv}{dt}=f.$$

(*e*) **Motion in a resisting medium.**

Suppose that a body is moving in a medium in which the velocity keeps on decreasing at a rate which is propotional to the velocity. If v be the velocity at time t, the rate of *decrease* of velocity at time t is $-\frac{dv}{dt}$. Therefore the motion of the body satisfies the differential equation

$$-\frac{dy}{dt}=kv,$$

where k is the constant of proportionality.

(*f*) **Exponential growth.**

Suppose that in a culture of yeast at each instant the time rate of change of active ferment $y(t)$ is proportional to the amount present. Then y satisfies the differential equation

$$\frac{dy}{dt}=ky,$$

where $k>0$ is the constant of proportionality.

(*g*) **Exponential decay.**

Experiments show that radium disintegrates at a rate proportional to the amount of radium present at any instant. If $y(t)$ denotes the amount of radium at time t, then

$$\frac{dy}{dt}=-ky.$$

where $k>0$ is the constant of proportionality. The negative sign on the right hand side indicates that y is decreasing as t is increasing.

Given an equation containing an arbitrary constant, it is possible to find a differential equation whose solution is the given equation. This is done by differentiating the given equation, and then eliminating the arbitrary constant from the given equation with the help of this equation.

Example 3. *Form the differential equation satisfied by* $y=(x+C)\,e^x$ *for all values of C.*

Solution. Differentiating both sides of the equation

$$y=(x+C)\,e^x \qquad \ldots(1)$$

with respect to x, we have

$$\frac{dy}{dx}=e^x+(x+C)\,e^x. \qquad \ldots(2)$$

Eliminating C from (1) and (2), we have

$$\frac{dy}{dx}=e^x+y,$$

as the desired differential equation.

EXERCISE 7 (*a*)

Find the order (O) and degree (D) of each of the following differential equations:

1. $\dfrac{dy}{dx}+3y=4x.$

2. $\dfrac{d^2y}{dx^2}+\left(\dfrac{dy}{dx}\right)^2-5y=e^x.$

3. $\left(\dfrac{d^3y}{dx^3}\right)^2+\left(\dfrac{dy}{dx}\right)^3-y=\sin x.$

4. $\left(\dfrac{d^4y}{dx^4}\right)^2+\left(\dfrac{d^3y}{dx^3}\right)^5-y=x^{10}.$

5. $\left(\dfrac{dy}{dx}\right)^2+y^2=\dfrac{d^2y}{dx^2}.$

6. Verifiy that $y=e^{2x}$ is a solution of the differential equation

$$\frac{dy}{dx}-2y=0.$$

In each of the problems 7—9 verify that the given function is a solution of the given differential equation.

7. $y''+y=0,\ y=A\cos x+B\sin x.$

8. $y''-y=4x,\ y=Ae^x+Be^{-x}-4x.$

9. $y'+2xy=0, y=ce^{-x^2}$.

10. Find a first-order differential equation involving both y and $\frac{dy}{dx}$ for which $y=5x^4$ is a solution.

11. Verify that $\frac{dy}{dx}+y=4$ is satisfied by $y=Ce^{-x}+4$. Find the value of C such that the resulting particular solution satisfies the condition $y=6$ when $x=0$.

12. Find the differential equation whose general solution is $y=C\cos x+x$.

13. Find the differential equation satisfied by the family of all parabolas $y^2=4ax$.

14. Find the differential equation of the family of all hyperbolas $xy=k$.

15. Find the differential equation of the family of all ellipses

$$\frac{x^2}{C+2}+\frac{y^2}{C}=1.$$

7·3. SOLUTION OF EQUATIONS BY THE METHOD OF SEPARATION OF VARIABLES

Some first-order differential equations can be reduced to the form

$$h(y)\frac{dy}{dx}=g(x) \qquad \cdots(1)$$

It is convenient to write (1) in the form

$$h(y)\,dy=g(x)\,dx \qquad \cdots(2)$$

An equation such as (1) is called an equation with separable variables (or a separable equation) because the variables can be separated in such a way that x appears only on one side and y appears only on the other side. By integrating both sides of (2), we have

$$\int h(y).\,dy=\int g(x)\,dx+C,$$

where C is the constant of integration.

By evaluating the integrals $\int h(y)\,dy$ and $\int g(x)\ dx$ we can find the general solution.

Example 4. *Find the general solution of the differential equation*

$$x^3\frac{dy}{dx}=y^2$$

Solution. The differential equation

$$x^3 \frac{dy}{dx} = y^2$$

can be written as

$$\frac{dy}{y^2} = \frac{dx}{x^3} \qquad \text{.. (1)}$$

Integrating both sides of (1), we have

$$\int \frac{dy}{y^2} = \int \frac{dx}{x^3} + C,$$

or

$$-\frac{1}{y} = -\frac{1}{2x^2} + C,$$

or

$$2x^2 - y + 2x^2 y\, C = 0, \qquad \text{...(2)}$$

where C is an arbitrary constant, is the desired solution.

Remark. Check that (2) is actually a solution of the given differential equation.

Example 5. *Solve* :

$$(x^2+1)\frac{dy}{dx} + y^2 + 1 = 0.$$

Solution. By separating the variables we can re-write the equation

$$(x^2+1)\frac{dy}{dx} + y^2 + 1 = 0 \qquad \text{...(1)}$$

as

$$\frac{dy}{y^2+1} = -\frac{dx}{x^2+1} \qquad \text{...(2)}$$

Integrating both sides of (2), we have

$$\int \frac{dy}{y^2+1} = -\int \frac{dx}{x^2+1} + C,$$

where C is an arbitrary constant,

or

$$\arctan y = -\arctan x + C$$

or

$$\arctan y + \arctan x = C, \qquad \text{...(3)}$$

which is the desired solution.

Remark. By taking tangents of both sides of (3), we have $\tan(\arctan y + \arctan x) = \tan C$

or

$$\frac{y+x}{1-y-x} = \tan C = C',$$

where we have written C′ in place of tan C.

The solution now takes the form

$$x+y=C'(1-xy)$$

or
$$y=\frac{C'-x}{1+C'x} \qquad \text{...(4)}$$

(4) is the general solution of (1) where C′ is an arbitrary constant. The two solutions (3) and (4) are equivalent. Either of the two is as good as the other. Furthermore, instead of C′ we can as well write C and write (4) as

$$y=\frac{C-x}{1+Cx}. \qquad \text{...(5)}$$

Remember that 'C' in (5) is *not* equal to 'C' in (3), even though both are arbitrary constants.

Sometimes a differential equation may not be in variables separable form as such, but it might reduce to that form by a simple substitution. The substitution is often suggested by the form of the differential equation.

Example 6. *Solve* :

$$(2x+y+1)\,dx+(4x+2y-1)\,dy=0.$$

Solution. Let $2x+y=v$. Then $2+\frac{dv}{dx}=\frac{dy}{dx}$, so that

$\frac{dy}{dx}=\frac{dv}{dx}-2$. The given equation becomes

$$(v+1)+(2v-1)\left(\frac{dv}{dx}-2\right)=0,$$

or
$$\frac{dv}{dx}=\frac{3v-3}{2v-1}.$$

Separating the variables and integrating we have

$$\int\frac{2v-1}{3v-3}\,dv=x+C,$$

or
$$\frac{2}{3}v+\frac{1}{3}\,ln\,|\,v-1\,|=x+C,$$

or
$$\frac{2}{3}(2x+y)+\frac{1}{3}\,lx\,|\,2x+y-1\,|=x+C.$$

or
$$\frac{1}{3}x+\frac{2}{3}y+\frac{1}{3}\,ln\,|\,2x+y-1\,|=C.$$

7·3·1. To obtain a particular solution of differential equation

In many applications we are not interested in the general solution of a given differential equation but only in the particular solution $y=y(x)$ satisfying a given condition to the effect that for

some value of x, say $x=x_0$, y has the value $y=y_0$, *i.e.*, $y(x_0)=y_0$. Such a condition is called an initial condition. A first order differential equation with an initial condition is called an *initial value problem*. In order to solve such a problem, we first solve the given differential equation, and then determine the value of the constant of integration by making use of the initial condition.

Example 7. *Solve :* $(y+2)\,y'=\sin x,\ y\,(0)=0$.

Solution. The differential equation

$$(y+2)\frac{dy}{dx}=\sin x, \qquad \text{...(1)}$$

can be written as

$$(y+2)\,dy=\sin x\,dx \qquad \text{...(2)}$$

Equation (2) is in 'variables separable' form. Integrating both sides we have

$$\int (y+2)\,dy=\int \sin x\,dx+C,$$

where C is an arbitrary constant.

or $$\frac{1}{2}y^2+2y=-\cos x+C,$$

or $$\frac{1}{2}y^2+2y+\cos x=C, \qquad \text{...(3)}$$

is the general solution.

We are given that $y(0)=0$, *i.e.*, when $x=0$, $y=0$. Let us find the particular solution satisfying this condition. Substituting $x=0$, $y=0$ in (3), we have

$$1=C.$$

When $C=1$, (3) reduces to

$$\frac{1}{2}y^2+2y+\cos x=1,$$

or $$y^2+4y+2\cos x-2=0,$$

which is the desired solution.

7·3·2. Some interesting applications

We shall now consider applications of differential equations to some geometrical and physical problems.

Example 8. *Find the curve through the point (3, 2) having at each of its points the slope $-x/4y$.*

Solution. The slope of the tangent to a curve at any point is $\dfrac{dy}{dx}$, which is given to be $-\dfrac{x}{4y}$. Therefore the differential equation of the family of curves satisfying the given condition is

$$\frac{dy}{dx}=\frac{-x}{4y}. \qquad \text{...(1)}$$

Writing (1) in 'variables separable' form, we have

$$4y\,dy=-x\,dx. \qquad \text{...(2)}$$

Integrating both sides of (2), we have

$$\int 4y\,dy=-\int x\,dx+C,$$

or
$$2y^2=-\frac{1}{2}x^2+C,$$

or
$$2y^2+\frac{1}{2}x^2=C, \qquad \text{...(3)}$$

where C is an arbitrary constant. (3) is the equation of the family of all curves satisfying (1). We have to find that particular curve of the family (3) which passes through the point (3, 2). Substituting $x=3$, $y=2$ in (3), we find that

$$2.2^2+\frac{1}{2}\cdot 3^2=C,$$

or
$$C=\frac{25}{2}$$

Substituting $C=\frac{25}{2}$ in (3), and simplifying, we have

$$x^2+4y^2=25,$$

as the equation of the desired curve.

Remark. Verify that $x^2+4y^2=25$ satisfies all the conditions stated in the problem.

Example 9. (*Newton's Law of Cooling*). *A copper ball is heated to a temperature of 100° and is then placed in water which is maintained at a temperature of 40°C. At the end of 5 minutes, the temperature of the ball is reduced to 60°. Find the time at which the temperature of the ball is reduced to 41°C.*

[*According to Newton's law of cooling, the rate of change of temperature T of the ball is proportional to the difference between T and the temperature of the surrounding medium.*

Assume that heat flows so rapidly in copper that at any time the temperature is the same at all points of the ball].

Solution.

Step 1. Let time t be measured from the instant when the ball is placed in water and let T (t) denote the temperature of the ball at time t. For the sake of brevity we shall write simply T

instead of T (t) but we are given that at $t=0$, T $(t)=100$, *i.e.*, T $(0)=100$. According to Newton's law of cooling, $\frac{dT}{dt}$ is proportional to T—40. If we denote the constant of proportionality by $-k$ (so that $k>0$), then we get the differential equation

$$\frac{dT}{dt}=-k\,(T-40). \qquad ...(1)$$

Equation (1) is the mathematical formulation of the law cooling in our case. We shall solve it to get T as a function of t.

Step 2. By separating the variables, (1) can be written as

$$\frac{dT}{T-40}=-k\,dt,$$

so that

$$T=Ce^{-kt}+40, \qquad ..(2)$$

where C is the constant of integration.

Step 3. The given initial condition is T $(0)=100$. Substituting $t=0$, $T=100$ in (2) we have

$$100=C+40, \text{ i.e., } C=60.$$

Substituting $C=60$ in (2), we find that the particular solution of (2) satisfying the given initial condition is

$$T=60\,e^{-kt}+40 \qquad ...(3)$$

Step 4. We shall now find k by using the information T $(5)=60$. Substituting $t=5$, $T=60$ in (3), we have

$$60=60\,e^{-5k}+40,$$

or

$$k=\tfrac{1}{5}\ln 5. \qquad ...(4)$$

Substituting the value of k in (3), we have

$$T=60\,e^{-(\frac{1}{5}\ln 5)t}+40, \qquad ...(5)$$

From (5) we find that $T=41$, when

$$1=60\,e^{-(k)t}$$

or

$$t=\frac{\ln 60}{\frac{1}{5}\ln 5}=\frac{3\log_{10} 60}{\log_{10} 5}=7{\cdot}6318=7{\cdot}63\ldots \text{ (approx.)}$$

Thus the temperature of the ball is reduced to 41° C after 7·63 minutes.

EXERCISE 7 (*b*)

Solve :

1. $y\frac{dy}{dx}=\cos x.$
2. $\frac{dy}{dx}=y^3.$
3. $\frac{1}{x}\frac{dy}{dx}=\frac{1}{x^2+3}.$
4. $\frac{dv}{du}=\frac{v+2}{u+1}.$

5. $\cos cx \dfrac{dy}{dx}=e^{y}$.

6. $\dfrac{r\,dr}{d\theta}=\cos 2\theta$.

7. $v^3\dfrac{dv}{dt}=(t+1)^2$.

8. $\cos^2 x \dfrac{dy}{dx}=2y^2 \cot x$.

9. $r\dfrac{d\theta}{dr}=\sec^2 \theta$.

10. $(x+4)\dfrac{dy}{dx}=y$.

11. $\dfrac{dy}{dx}=e^{x+y}$ (*D.B S.S.C.E., 1984*)

12. $(e^x+1)\, y\, dy=(y+1)\, e^x\, dx$ (*D.B.S S.C.E., 1988*)

13. $(1-y)\, x \dfrac{dy}{dx}+(1+x)\, dy=0$ (*A.I.S.S.C.E., 1984*)

14. $(1+\cos x)\, dy=(1-\cos x)\, dx$. (*A.I.S S.C.E., 1984*)

15. $\csc x \ln y\, dy+x^2 y\, dx=0$. (*A.I S,S.C.E., 1986*)

16. $\cos x \;\; \cos y \dfrac{dy}{dx}=-\sin x \sin y$ (*D.B.S.S.C.E., 1987*)

17. $x \cos y\, dy=(xe^x \ln x+e^x)\, dx$. (*D.B.S.S.C.E., 1988*)

18. $\sqrt{a+x}\dfrac{dy}{dx}=-xy$. (*A.I.S.S C.E., 1984*)

19. $(1-x^2)\, dy+xy\, dx=xy^2\, dx$. (*D.B.S.S.C.E., 1989*)

20. Find the particular solution of $xy'+y=0$, satisfying the condition $y=1$ when $x=1$.

21. Find the particular solution of

$$xy\frac{dy}{dx}=y+3$$

satisfying the condition $y=0$ when $x=1$.

22. Solve $y'=y \tan x$ subject to $y(\pi)=-1$.

23. (**Exponential growth**). In a culture of yeast the time rate of change of active ferment $y(t)$ at any instant of time t is proportional to the amount present at that instant. If $y(t)$ doubles in 3 hours, how much should we expect at the end of 9 hours at the same rate of growth ?

24. (**Exponential decay**). Experiments show that radium disintegrates at a rate proportional to the amount of radium present at any instant. If its half-life, that is, the time in which 50% of a given amount will disintegrate, be known to be 1590 years, what percent will disintegrate in one year ?

25. (**Newton's Law of Cooling**). Experiments show that the rate of change of temperature T of a body is proportional to the difference between T and the temperature of the surrounding medium.

A copper ball is heated to a temperature of 100°C. It is placed in water which is maintained at a temperature of 20°C. At the end of 5 minutes the temperature of the ball is reduced to 60° C. Find the time at which the temperature of the ball is reduced to 21° C.

26. A wet porous substance when exposed to open air loses its moisture at a rate proportional to the moisture content. A sheet of the substance exposed to wind loses half its moisture during one hour. In how much time would it have lost 99% of the moisture, weather conditions remaining the same throughout ?

27. Two liquids are boiling in a vessel. If the ratio of the quantities of each passing off as vapour at any instant is proportional to the quantities x and y still in the liquid state, show that $x\dfrac{dy}{dx}-ky=0$, where k is the constant of proportionality, and determine y in terms of x.

28. Find the curve in the xy-plane passing through the point $(9, 3)$ and having at each of its points the slope x^2.

29. Find the curve in the xy-plane passing through the point $(-3, 1)$ and having slope $\dfrac{y+1}{x^2-1}$.

30. Find the curve in the xy-plane all normals of which pass through the point $(1, 0)$.

7·4. HOMOGENEOUS EQUATIONS AND THEIR SOLUTION

A function $f(x, y)$ of two variables x and y is said to be *homogeneous* in x and y if it can be expressed in the form $x^n\, g(y/x)$ for some function g. Furthermore, we then say that f is of degree n.

Consider the following examples :

(*a*) $f(x, y)=x^2y+xy^2+x^3$.

Since $f(x, y)=x^3\,[y/x+(y/x)^2+1]$,

therefore f is a homogeneous function of x and y of degree 3.

(*b*) $f(x, y)=x^3+y^3+x+y$.

f is not a homogeneous function of x and y.

(*c*) $f(x, y)=\sin(y/x)$ is a homogeneous function of x and y of degree zero because we can write it as

$$x^0\, g(y/x), \text{ where } g(t)=\sin t.$$

(*d*) $f(x, y)=y^2 \arctan(y/x)$ is a homogeneous function of x and y of degree 2 (why ?)

Consider the following differential equations :

(A) $x(x-y)\dfrac{dy}{dx}+y^2=0$.

It is of the form $M\frac{dy}{dx}+N=0$,

where $M=x(x-y)$, $N=y^2$. M and N are both homogeneous functions of x and y of degree 2.

(B) $x\frac{dy}{dx}=y-x\tan\frac{y}{x}$.

It is of the form

$M\frac{dy}{dx}+N=0$, where

$M=x$, $N=x\tan(y/x)-y$. M and N are both homogeneous functions of x and y of degree 1.

A first-order differential equation is said to be *homogenous* if it can be expressed in the form

$$M\frac{dy}{dx}+N=0,$$

where M and N are homogeneous functions of x and y of the same degree.

The differential equations (A) and (B) above are homogeneous differential equations.

Neither of the following differential equations is homogeneous.

(C) $(2x+3y-5)\frac{dy}{dx}+(3x+2y-1)=0$.

Observe that it cannot be expressed in the form $M\frac{dy}{dx}+N=0$ where M, N are homogeneous functions.

(D) $(2x^2y+3xy^2)\frac{dy}{dx}+x^2-y^2=0$.

This equation is not a homogeneous equation even though $2x^2y+3xy^2$ and x^2-y^2 are both homogeneous functions.

(Reason : The two functions are not of the same degree)

7·4·1. Solution of a homogeneous differential equation of the first order.

$$\text{Let } M\frac{dy}{dx}+N=0 \qquad \cdots(1)$$

be a homogeneous differential equation, where M and N are homogeneous functions of degree N. Let us write

$$M=x^n g(y/x),\ N=x^n h(y/x). \qquad \cdots(2)$$

Using (2), we can re-write (1) as

$$\frac{dy}{dx}=\frac{-N}{M}=\frac{-x^n\ h\ (y/x)}{x^n g\ (y/x)}=\frac{-h\ (y/x)}{g\ (y/x)}=f\ (y/x),$$

where $f\ (y/x)$ is some function of y/x.

Thus we find that (1) can be expressed in the form

$$\frac{dy}{dx}=f\ (y/x),$$

where f is a given function of y/x.

The form of the function f suggests that we ought to put $y/x=v$.

Remembering that y is a function of x (to be determined), v is also a function of x. Thus, letting $y=vx$, we have by differentiation,

$$\frac{dy}{dx}=v+x\ \frac{dv}{dx}=f\ (v).$$

By separating the variables, we find

$$\frac{dv}{f\ (v)-v}=\frac{dx}{x} \qquad \cdots(3)$$

Integrating (3) and replacing v by y/x we shall have the general solution of (1). The following examples will illustrate the method.

Example 10. *Solve* :

$$x\ (x-y)\frac{dy}{dx}+y^2=0. \qquad \cdots(1)$$

(*Roorkee Entrance, 1980*)

Solution. The given equation is a homogeneous differential equation. Therefore we shall use the substitution $y=vx$. ...(2)

Differentiating $y=vx$ with respect to x, we have

$$\frac{dy}{dx}=v+x\frac{dv}{dx} \qquad \cdots(3)$$

Substituting the values of y and $\frac{dy}{dx}$ in (1), we have

$$x\ (x-vx)\ \left(v+x\frac{dv}{dx}\right)+v^2x^2=0,$$

or

$$(1-v)\left(v+x\frac{dv}{dx}\right)+v^2=0,$$

or

$$v\ (1-v)+(1-v)x\ \frac{dv}{dx}+v^2=0,$$

or

$$v+(1-v)\ x\frac{dv}{dx}=0.$$

Separating the variables, we have

$$\left(1-\frac{1}{v}\right)\ dv=\frac{dx}{x}$$

Integrating both sides, we have

$$\int\left(1-\frac{1}{v}\right) dv=\int\frac{dx}{x}+c,$$

or $$v-\ln|v|=\ln|x|+c,$$

or $$v=\ln|vx|+c \qquad ...(4)$$

Substituting $v=y/x$ in (4), we have

$$y/x=\ln|y|+c,$$

or $$e^{y/x}=|y|\,e^c,$$

or $$|y|=e^{-c}\,e^{y/x}=\mathrm{C}\,e^{y/x},$$

Thus $$|y|=\mathrm{C}e^{y/x}$$

is the desired solution, where C is an arbitrary constant.

Example 11. *A curve passes through the point $(1, \pi/4)$ and its slope at any point (x, y) on it is*

$$\frac{dy}{dx}=\frac{y}{x}+\cos^2\frac{y}{x}$$

Determine the curve. (*Roorkee Entrance, 1981*)

Solution. Since the equation

$$\frac{dy}{dx}=\frac{y}{x}+\cos^2\frac{y}{x} \qquad ...(1)$$

is homogeneous in x and y, therefore let us use the substitution

$$x=vx \qquad ...(2)$$

Differentiating (2) throughout w.r.t. x, we have

$$\frac{dy}{dx}=v+x\frac{dv}{dx} \qquad ...(3)$$

On using (2) and (3), (1) reduces to

$$x\frac{dv}{dx}=\cos^2 v,$$

Separating the variables, we have

$$\sec^2 v\,dv=\frac{dx}{x}.$$

Integrating both sides, we have

$$\int\sec^2 v\,dv=\int\frac{dx}{x}+\mathrm{C},$$

or $$\tan v=\ln|x|+\mathrm{C}.$$

or $$\tan(y/x)=\ln|x|+\mathrm{C}. \qquad ...(4)$$

Since we are given that $y=\pi/4$ when $x=1$, therefore from (4), we have

$$\tan\pi/4=\ln 1+\mathrm{C}.$$

or $$\mathrm{C}=1.$$

Substituting the value of C in (4), we have

$$\tan(y/x)=\ln|x|+1$$

as the desired equation.

EXERCISE 7 (*c*)

1. $(x+y)\,dx+x\,dy=0.$
2. $(x-y)\,dx+(x+y)\,dy=0.$
3. $(x+y)\,dx+(y-x)\,dy=0.$
4. $(x^2+y^2)\,dx-2xy\,dy=0.$
5. $2x^2\dfrac{dy}{dx}=xy+y^2.$
6. $(x^2+y^2)\dfrac{dy}{dx}=xy.$
7. $y^2+x^2\dfrac{dy}{dx}=xy\dfrac{dy}{dx}.$
8. $x^2y\,dx-(x^3+y^3)\,dy=0.$
9. $x\sin(y/x)\,dy=[y\sin(y/x)-x]\,dx.$
10. $x\dfrac{dy}{dx}=y-x\tan(y/x).$ (*Roorkee Entrance, 1982*)
11. Find the particular solution of the differential equation
$$(x+y)\,dx+(x-y)\,dy=0$$
such that $y=1$ when $x=1$.

7·5· LINEAR DIFFERENTIAL EQUATIONS OF THE FIRST ORDER

A differential equation of the first order is said to be *linear* if it can be written in the form

$$\frac{dy}{dx}+Py=Q. \qquad ...(1)$$

where P and Q may be functions of x (or constants) but do not depend on y. For example, the following equations are all linear :

(*a*) $\dfrac{dy}{dx}+xy=x^2,$

(*b*) $\dfrac{dy}{dx}+4y=\sin x,$

(*c*) $\dfrac{dy}{dx}+xy=0,$

(*d*) $\sin x\dfrac{dy}{dx}-\cos x=y\cot x.$

None of the following equations is linear :

(*e*) $\left(\dfrac{dy}{dx}\right)^2+4y=e^x,$

(*f*) $y\dfrac{dy}{dx}+2x=y^2$,

(g) $\dfrac{dy}{dx}+ye^x=y^3$.

Equation (1) is said to be linear (in y) beaause it is linear in y and $\dfrac{dy}{dx}$. Products and powers (higher than the first) of y and $\dfrac{dy}{dx}$ do not occur in it. This property characterizes a linear differential equation.

If in (1), Q=0, then we say that it is a homogeneous equation ; otherwise we say that it is a nonhomogeneous equation. In the illustrations given above, equation (*c*) is homogeneous, equations (*a*), (*b*) and (*d*) are all nonhomogeneous.

If P is a constant, then we say that equation (1) is a linear equation with constant coefficients. Equation (*b*) in the examples given above is a linear differential equation of the first order with constant coefficients. The solution of such an equation can be obtained easily as follows :

Let $$\frac{dy}{dx}+Py=Q \qquad ...(2)$$

be a linear differential equation of the first order with constant coefficients, so that P is a constant, and Q is a function of x.

To solve (2), let us first consider the corresponding homogeneous equation

$$\frac{dy}{dx}+Py=0. \qquad ...(3)$$

By separating the variables it can be seen that the solution of (3) is

$$ye^{Px}=C,$$

where C is arbitrary constant. This *suggests* that the solution of (2) is somewhat similar in form, and it might be possible to solve (2) by the substitution

$$ye^{Px}=v \qquad ...(4)$$

Differentiating (4) throughout with respect to x, we have

$$\frac{dy}{dx}e^{Px}+Pye^{Px}=\frac{dv}{dx}$$

or $$\left(\frac{dy}{dx}+Py\right)e^{Px}=\frac{dv}{dx}$$

or $$\frac{dv}{dx}=Qe^{Px}, \text{ by using (2).}$$

Integrating with respect to x, we have

$$v=\int Qe^{Px}\,dx+C,$$

or
$$\boxed{e^{Px} = Qe^{Px}\, dx + C} \qquad ...(4)$$

is the desired general solution of (2), where C is an arbitrary constant.

Remark. e^{Px} is generally called the integrating factor for the differential equation (2) and is denoted by I.F.

To solve an equation such as (2), we do not usually work out all the details as indicated above, but simply use (5). The following examples will illustrate the procedure.

Example 12. *Solve :*

$$\frac{dy}{dx} - 4y = e^{x}.$$

Solution. The given equation is a linear differential equation of the first order with constant coefficients.

I.F. $= e^{Px} = e^{-4x}$, since $P = -4$.

The general solution is

$$ye^{Px} = \int Qe^{Px}\, dx + C,$$

i.e.,
$$ye^{-4x} = \int e^{x} \cdot e^{-4x}\, dx + C,$$

or
$$ye^{-4x} = \int e^{-3x}\, dx + C = -\frac{1}{3}e^{-x} + C,$$

or
$$y = -\tfrac{1}{3}e^{3x} + Ce^{4x},$$

where C is an arbitrary constant.

Example 13. *Solve :*

$$\frac{dy}{dx} + 2y = \cos 3x.$$

Solution. The given differential equation is a linear differential equation of the first order with constant coefficients.

I.F. $= e^{2x}$.

The general solution is

$$ye^{2x} = \int e^{2x} \cos 3x\, dx + C \qquad ...(1)$$

Let
$$I = \int e^{2x} \cos 3x\, dx.$$

By integrating by parts taking e^{2x} as the first function and $\cos 2x$ as the second function, we have

$$I = e^{2x}\frac{(\sin 3x)}{3} - \int 2e^{2x}\frac{(\sin 3x)}{3}\,dx,$$

$$= \frac{1}{3}e^{2x}\sin 3x - \frac{2}{3}\int e^{2x}\sin 3x\,dx,$$

$$= \frac{1}{3}e^{2x}\sin 3x - \frac{2}{3}\left[e^{2x}\frac{\cos 3x}{-3} - \int 2e^{2x}\frac{\cos 3x}{-3}\,dx\right],$$

$$= \frac{1}{9}e^{2x}(-3\sin 3x + 2\cos 3x) - \frac{4}{9}I,$$

so that $I = \frac{-3}{13}e^{2x}\sin 3x + \frac{2}{13}e^{2x}\cos 3x.$ ...(2)

Substituting the value of I from (2) in (1), we have

$$ye^{2x} = \frac{1}{13}e^{2x}(-3\sin 3x + 2\cos 3x) + C,$$

or
$$y = \frac{1}{13}(-3\sin 3x + 2\cos 3x) + Ce^{-2x}$$

as the general solution, C being an arbitrary constant.

Example 14. *Solve :*

$$\frac{dy}{dx} - 3y = x.$$

Solution. The given differential equation is a linear differential equation of the first order with constant coefficients,

I.F. $= e^{-3x}$

The general solution is

$$ye^{-3x} = \int xe^{-3x}\,dx + C,$$

$$= x\left(-\frac{1}{3}e^{-3x}\right) - \int 1\cdot\left(-\frac{1}{3}e^{-3x}\right)dx + C,$$

$$= -\frac{1}{3}xe^{-3x} - \frac{1}{9}e^{-3x} + C,$$

or
$$y = -\frac{1}{3}x - \frac{1}{9} + Ce^{3x},$$

where C is an arbitrary constant.

EXERCISE 7 (*d*)

Solve :

1. $\frac{dy}{dx} + y = 1.$
2. $\frac{dy}{dx} - y = x - 1.$

3. $\frac{dy}{dx}-4y=4x^2-2x.$

4. $\frac{dy}{dx}+3y=e^{2x}+12.$

5. $\frac{dy}{dx}-3y=e^{3x}.$

6. $\frac{dy}{dx}-y=e^{x}.$

7. $\frac{dy}{dx}-3y=13 \sin 2x.$

8. $\frac{dy}{dx}+2y=5 \cos x.$

9. $\frac{dy}{dx}-4y=e^{4x} \sin 2x.$

10. $\frac{dy}{dx}+3y=2e^{-2x} \cos x.$

TEST YOUR UNDERSTANDING VII

In each of the following problems, four alternatives are given out of which exactly one is correct. Put a tick mark (✓) against the correct alternative.

1. The differential equation

$$4 \frac{d^2y}{dx^2}+3\left(\frac{dy}{dx}\right)^2+y=0$$

is of order

(a) 1 (b) 2 (c) 3 (d) 4.

2. The degree of the differential equation

$$4\left(\frac{d^2y}{dx^2}\right)^3+\left(\frac{dy}{dx}\right)^5+y^7=x^9 \text{ is}$$

(a) 9 (b) 7 (c) 5 (d) 3.

3. The differential equation

$$5\left(\frac{d^2y}{dx^2}\right)^2+\left(\frac{dy}{dx}\right)^4-xy=0 \text{ is of}$$

(a) order 2 and degree 4 (b) order 4 and degree 2
(c) order 2 and degree 2 (d) order 5 and degree 1.

4. The general solution of the equation

$$xdy-ydx=0 \text{ is}$$

(a) $xy=1$ (b) $xy=C$
(c) $y=x$ (d) $y=Cx$.

5. A particular solution of the equation

$$y \frac{dy}{dx}-x=0 \text{ is}$$

(a) $y^2+x^2=1$ (b) $y^2-x^2=1$
(c) $y^2+x^2=C$ (d) $xy=1$.

6. Which of the following equations is of variables separable form ?

(a) $\sin(x+y)\dfrac{dy}{dx}=1$, (b) $xy\dfrac{dy}{dx}=x^2+1$,

(c) $\dfrac{dy}{dx}=\cos(y/x)$, (d) $xy\dfrac{dy}{dx}=x^2+y^2$.

7. Which one of the following equations is homogeneous :

(a) $(2x+y)\dfrac{dy}{dx}=x-2$, (b) $x^2\dfrac{dy}{dx}=y^2-x^2$,

(c) $\dfrac{dy}{dx}=e^{x+y}$, (d) $x^2\,dy+(y^2-2)\,dx=0$.

8. Which one of the following equations is linear :

(a) $\dfrac{dy}{dx}+xy=1$, (b) $x^2\dfrac{dy}{dx}+y=e^x$,

(c) $\dfrac{dy}{dx}+3y=xy^2$, (d) $x\dfrac{dy}{dx}+y^2=\sin x$.

9. Which one of the following equations is a linear equation with constant coefficients :

(a) $x\dfrac{dy}{dx}-\sin x=xy$, (b) $\dfrac{dy}{dx}-xy=4$,

(c) $y\dfrac{dy}{dx}+x^3=\cos x$, (d) $\dfrac{dy}{dx}+2y^2=e^x$.

10. The general solution of the equation $\dfrac{dy}{dx}-y=e^x$ is

(a) $ye^x=\frac{1}{2}e^{2x}+C$, (b) $ye^x=x+C$,

(c) $y=(C-x)e^x$, (d) $y=(x+C)e^x$.

REVIEW EXERCISE VII

Solve :

1. $\dfrac{dy}{dx}=1+x+y+xy$. *(A.I.S.S.C.E., 1985)*
2. $(x^2y-x^2)\,dx+(xy^2-y^2)\,dy=0$. *(D.B.S.S.C.E., 1984)*
3. $3e^x\tan y+(1-e^x)\sec^2 y\dfrac{dy}{dx}=0$. *(A.I.S.S.C.E., 1986)*
4. $(1+x)(1+y^2)\,dx+(1+y)(1+x^2)\,dy=0$. *(D.B.S.S.C.E., 1986)*
5. $\sqrt{a+x}\,\dfrac{dy}{dx}=-xy$. *(A.I.S.S.C.E., 1984)*
6. $(4x+6y+5)\,dx=(2x+3y+4)\,dy$.
7. $\dfrac{dy}{dx}=1+(y-x)\cot x$.

8. $\frac{dy}{dx}+\frac{x-2y}{2x-y}=0.$

9. $(y^3-2yx^2)\,dx+(2xy^2-x^3)\,dy=0.$

10. $(x^2+2xy-y^2)\,dx+(y^2+2xy-x^2)\,dy=0.$

11. $\frac{dy}{dx}+4y=2x^2+x.$

12. $\frac{dy}{dx}-2y=e^{2x}\cos 2x.$

13. $\frac{dy}{dx}+5y=2xe^{-5x}.$

14. $\frac{dy}{dx}-3y=e^{3x}+6\cos 3x.$

15. Find the solution of the differential equation

$$y-x\frac{dy}{dx}=a\left(y^2+x^2\frac{dy}{dx}\right)$$

satisfying $x=a, y=a.$ *(A.I.S.S.C.E., 1989)*

16. A car moves along a straight road in such a way that its velocity at t seconds is given by $S'(t)=3t^{1/2}$. How far does the car travel during the first 100 seconds ? *(Roorkee Entrance, 1981)*

17. Solve $x(1+y^2)\,dx+(1+x^2)\,dy=0$, given that $y=1$ when $x=0$. *(A.I.S S.C.E., 1986)*

18. Find the curve whose slope is

$$\frac{dy}{dx}=\frac{2y}{x},\ x>0,\ y>0$$

and which passes through the point (1, 1). *(Roorkee Entrance, 1981)*

SUMMARY

1. The *order* of a differential equation is the order of the highest derivative appearing in the equation.
2. The *degree* of a differential equation is the degree of the derivative of the highest order occurring in it, after expressing it in a form free from radicals so far as derivatives are concerned.
3. A differential equation of the first order and first degree is said to be in *variables separable form* if it can be expressed as $f(x)\,dx+g(y)\,dy=0$.
4. A differential equation of the first order and first degree is said to be *homogeneous* if it can be expressed in the form

$$\frac{dy}{dx}=f(y/x).$$

5. A homogeneous differential equation of the first order can be solved by the substitution $y=vx$.

6. A differential equation of the first order and first degree is said to be a *linear equation with constants coefficients* if it can be expressed in the form

$$\frac{dy}{dx}+Py=Q,$$

where P is a constant and Q is a function of x.

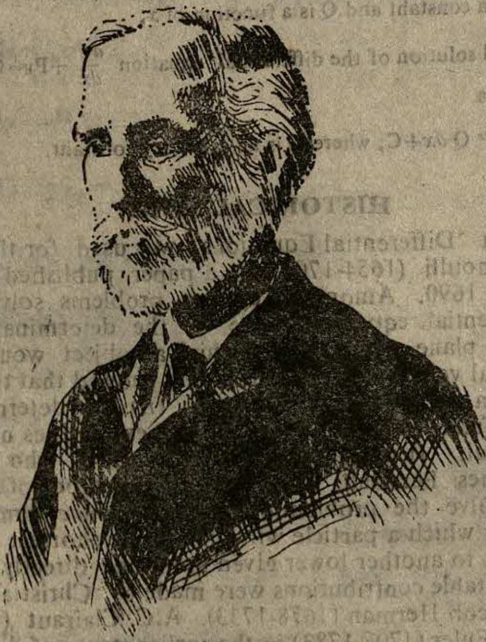
7. The general solution of the differential equation $\frac{dy}{dx}+Py=Q$, where P is a constant is

$$ye^{Px}=\int e^{Px}\,Q\,dx+C, \text{ where C is an arbitrary constant.}$$

HISTORICAL NOTE

The term 'Differential Equations' was used for the first time by James Bernoulli (1654-1705) in a paper published in the *Acta Eruditorum* for 1690. Among the earliest problems solved with the help of differential equations were (*i*) the determination of the isochrone—the plane curve along which an object would fall with uniform vertical velocity. James Bernoulli showed that the required curve must be a semi-cubical parabola ; (*ii*) the determination of the slope of a freely hanging string, the two extremities not being in the same vertical. It was James Bernoulli again, who solved this problem. James Bernoulli also used the method of differential equations to solve the problem of *Brachistochrone*, namely to find the curve along which a particle will slide in the shortest time from one given point to another lower given point not directly below the first point. Notable contributions were made by Christian Huygens (1675-1726), Jacob Herman (1678-1733), A.C. Clairaut (1713-1765) and Leonhard Euler (1707-1783) in the earlier stages of development of this important branch of Mathematics.

□ □



JOSIAH WILLARD GIBBS (1939—1903)

Josiah Willard Gibbs was born in New Haven, Connecticut, U.S.A. in 1839. His father was Professor of Sacred Literature at the University of Yale. Gibbs entered Yale in 1854 and graduated at the age of 19. Immediately thereafter he began graduate work in Engineering and obtained his Ph. D. at Yale in 1863, with a thesis on gear design. In the same year, he was appointed Tutor in Yale College. In 1871 he became Professor of Mathematical Physics. Gibbs' interest in vectors was aroused by his study of Maxwell's classical work 'Electricity and Magnetism'. In 1881 he published a pamphelet 'Elements of Vector Analysis' which was circulated privately. In 1884 he enlarged this pamphlet. Talking of Gibbs, Max Planck once said, "His name, not only in America, but in the whole world, will ever be reckoned among the most renowned theoretical physicists of all times." Gibbs led a quiet life. He worked with great dedication and passed away in 1903 in New Haven.

CHAPTER 8

Vectors

8·1. SCALARS AND VECTORS

In our daily lives we often come across quantities such as mass, volume, density, temperature, height, etc. which have magnitudes but are not related to any direction in space Such quantities are called *scalars*. To specify a scalar we require a unit quantity of the same type, and the ratio (k) which the given quantity bears to this unit. This enables us to express the quantity ask times the unit. The number k is called the measure of the quantity in terms of the selected unit. The measure is a real number, and real numbers thus suffice to express scalar quantities.

We also often come across quantities such as displacement, velocity, acceleration, force, momentum etc., which have not only magnitude, but which are also related to a definite direction, and which are such that two quantities of the same kind are compounded according to the triangle law of addition (to be stated later on). Such quantities are called *vectors*.

While ordinary algebra suffices to handle vector quantities as well, the treatment often becomes involved, and it is more convenient to develop a mathematical system which bears to vectors the same relation as the algebra of real numbers bears to scalars.

8·2. DIRECTED LINE SEGMENTS

To develop an adequate mathematical system which will suffice for our needs as indicated above, we shall introduce the notion of a directed line segment.

A directed line segment is the simplest type of entity having both magnitude and direction. It is determined by two points, of which one of the points is called the *initial point*, and the other is called the *terminal point*. The magnitude of the directed line segment is its length, and its direction is that from the initial point to the terminal point. It is denoted by an arrowhead. Thus for example, in Fig. 8·1, the points O and A determine a directed line segment, to be denoted by $\overrightarrow{OA}$, O being the initial point and A the terminal point. The magnitude of $\overrightarrow{OA}$ is the length OA, and its direction is from O to A. Two directed line segments are said to be equal if they have the same magnitude

Fig. 8·1.

and the same direction. Thus in Fig. 8·2, $\vec{AB}=\vec{CD}=\vec{EF}$, because $\vec{AB}$, $\vec{CD}$, $\vec{EF}$ all have the same magnitude, and also have the same direction. But $\vec{PQ}\neq\vec{EF}$ because the lengths PQ and EF are different, even though $\vec{PQ}$ and $\vec{EF}$ have the same direction.

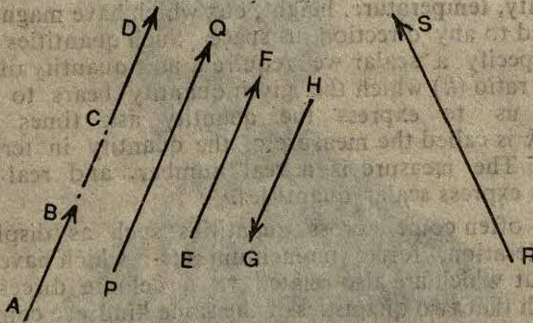


Fig. 8·2.

Again, $\vec{PQ}\neq\vec{RS}$ even though PQ=RS, because $\vec{PQ}$ and $\vec{RS}$ have different directions. The directed line segments $\vec{EF}$ and $\vec{GH}$ are equal in magnitude (EF=GH) and are parallel (EF ∥ GH) but are not equal, because their directions are opposite. Two such directed line segments are said to be the negatives of each other. For $\vec{EF}$ and $\vec{GH}$ we would thus write, $\vec{EF}=-\vec{GH}$ From the above discussion we find that we associate three attributes with a given directed line segment, namely, its length, support and sense The suport and sense both taken together specify the direction of the line segment.

(*a*) **Length.** The length of $\vec{PQ}$ is denoted by $|\vec{PQ}|$. It is obvious that $|\vec{PQ}|=|\vec{QP}|$.

(*b*) **Support**. The line of unlimited length of which a given directed line segment is a part, is called its support.

(*c*) **Sense**. The sense of $\vec{PQ}$ is from P to Q, and that of $\vec{QP}$ is from Q to P. Thus the directed line segments $\vec{PQ}$ and $\vec{QP}$ have opposite senses, even though they have the same magnitude and the same support. If two vectors have the same support and same sense

we often say that have the same direction. If two vectors have the same support but opposite senses, we often say that they have opposite directions.

8 3. VECTORS AS DIRECTED LINE SEGMENTS

A directed line segment can be used to represent a vector quantity in magnitude as well as in direction. To represent a vector quantity in this manner, we first choose a unit and express the magnitude of the vector quantity in terms of that unit Any directed line segment having the same direction as that of the vector quantity (to be represented) and magnitude equal to the measure of the vector quantity (in the unit selected) then represents the vector quantity.

Definition 8·1. *A directed line segment is called a* ***vector.*** In Fig. 8·2, $\vec{AB}$, $\vec{CD}$, $\vec{EF}$, $\vec{GH}$, $\vec{PQ}$, $\vec{RS}$ are all vectors.

8·3·1. Equality of two Vectors

Definition 8·2. *Two vectors are said to be equal if they have* (*i*) *the same length* (*ii*) *the same or parallel supports, and* (*iii*) *the same sense.*

Illustration. In Fig. 8·2, the vectors $\vec{AB}$, $\vec{CD}$ and $\vec{EF}$ are all equal to each other, but $\vec{PQ} \neq \vec{EF}$. From the above definition we find that two different line segments may stand for the same vector.

8·3·2. Notation for Vectors

We shall denote a vector by a single Clarendon (bold-faced) symbol such $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ etc. The magnitude of $\mathbf{a}$ will be denoted by $|\mathbf{a}|$. The fact of equality of two vectors $\mathbf{a}$ and $\mathbf{b}$ will be expressed in symbols by writing $\mathbf{a}=\mathbf{b}$.

8·3·3. Unit Vector

A vector whose magnitude is 1 is called a unit vector.

8·3·4. Zero Vector

A vector whose initial and terminal points coincide is called the zero vector. The length of the zero vector is zero, but we can think of it as having any line as its line of support. The zero vector will be denoted throughout by the symbol O. As we shall see later, the zero vector has many properties similar to those of the real number 0.

8·3·5. Co-initial Vectors

Our definition of equality of vectors enables us to replace a given vector by an equal vector but having any given point as initial point.

Let $\overrightarrow{AB}$ be any given vector, and let C be a given point. Two different cases arise, according as C lies on the support of $\overrightarrow{AB}$ or it dot not lie on the support of $\overrightarrow{AB}$.

If C lies on the support of $\overrightarrow{AB}$, choose a point D on it so that AB=CD, and the vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ have the same sense.

Then $\overrightarrow{AB}=\overrightarrow{CD}$.

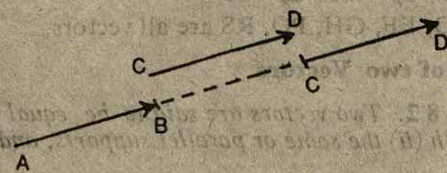


Fig. 8·3.

If C does not lie on the support of $\overrightarrow{AB}$, then we draw a line through C parallel to AB, and choose a point D on it so that $\overrightarrow{CD}$ has the same magnitude and sense as $\overrightarrow{AB}$. Then $\overrightarrow{AB}=\overrightarrow{CD}$.

Vectors with the same initial point are often called *co-initial vectors*. In view of the above discussion, given a set of vectors, we can always take all of them to be co-initial.

8 4. ADDITION OF VECTORS

We shall now how s:e two vectors **a** and **b** can be added to give rise to another vector and study some important properties of this operation of addition of vectors.

Definition 8·3. (*Triangle law of addition*). *Let* **a** *and* **b** *be two vectors ; let O be any point in space, and let* $\overrightarrow{OP}$ *represent* **a**. *If* $\overrightarrow{PQ}$ *represents* **b**, *then* $\overrightarrow{OQ}$ *represents* **a**+**b**.

Remark. From the above definition it appears as if the sum of the vectors **a** and **b** depends upon the choice of the point O. For the definition to be meaningful this should not be so. In fact, it can

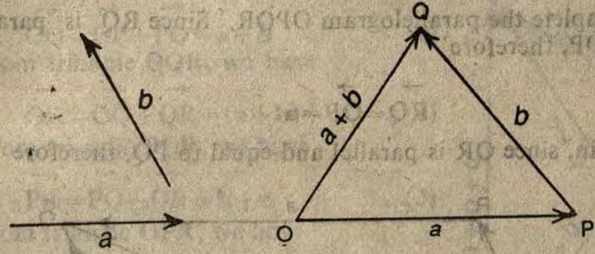


Fig. 8·4.

be easily seen (as follows) that in whatever manner we may choose point O, we arrive at the same vector :

Let O and O′ be any two points, and let

$$\overrightarrow{OP}=\mathbf{a}=\overrightarrow{O'P'} \qquad ...(1)$$

$$\overrightarrow{PQ}=\mathbf{b}=\overrightarrow{P'Q'}. \qquad ...(2)$$

From (1), OP and O′P′ are parallel and equal, so that O′P′PO is a ‖ gm, and consequently O′O and P′P are equal and parallel. From (2), PQ and P′Q′ are parallel and equal, so that P′PQQ′ is a ‖ gm, and consequently, P′P and Q′Q are equal and parallel.

Since P′P′ is parallel and equal to both O′O and Q′Q, therefore Q′Q is equal and parallel to O′O. It follows that Q′QOO′ is a ‖ gm, and therefore OQ is equal and parallel to O′Q′. Thus $\overrightarrow{PQ}=\overrightarrow{O'Q'}$. This shows that *the sum of two vectors is independent of the choice of the point O.*

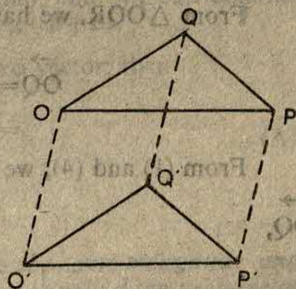


Fig. 8·5.

Theorem 8·1. *Addition of vectors is commutative, that is,*

$$\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$$

or any pair of vectors **a** *and* **b**.

Proof. Let **a** and **b** be any two vectors.

Let $\overrightarrow{OP}$ represent **a** and let $\overrightarrow{PQ}$ represent **b**.

By definition of addition of vectors, from △OPQ, we have

$$\overrightarrow{OQ}=\mathbf{a}+\mathbf{b} \qquad ...(1)$$

Complete the parallelogram OPQR. Since RQ is parallel and equal to OP, therefore

$$\vec{RQ}=\vec{OP}=\mathbf{a} \quad \text{...(2)}$$

Again, since OR is parallel and equal to PQ, therefore

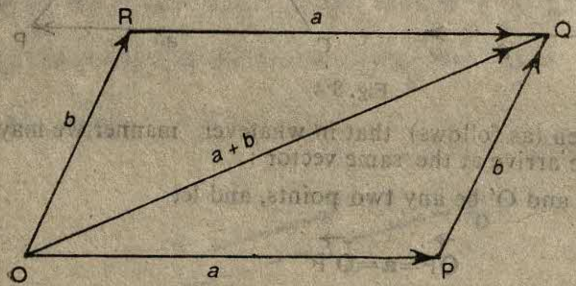


Fig. 8·6.

$$\vec{OR}=\vec{PQ}=\mathbf{b}. \quad \text{...(3)}$$

From $\triangle$OQR, we have

$$\vec{OQ}=\vec{OR}+\vec{RQ},$$

$$=\mathbf{b}+\mathbf{a} \quad \text{...(4)}$$

From (1) and (4), we have by equating the two expressions for $\vec{OQ}$,

$$\boxed{\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}}$$

Theorem 8·2. *Addition of vectors is associative, that is,*

$$\mathbf{a}+(\mathbf{b}+\mathbf{c})=(\mathbf{a}+\mathbf{b})+\mathbf{c}$$

for any three vectors **a**, **b** *and* **c**.

Proof. Let **a**, **b** and **c** be any three vectors. Let $\vec{OP}$ represent **a**, $\vec{PQ}$ represent **b**, and $\vec{QR}$ represent **c**.

By definition of addition of vectors, $\vec{OQ}=\mathbf{a}+\mathbf{b}$.

From triangle OPQ, we have

$$\vec{OQ}=\vec{OP}+\vec{PQ}=\mathbf{a}+\mathbf{b} \quad \ldots(1)$$

From triangle OQR, we have

$$\vec{OR}=\vec{OQ}+\vec{QR}=(\mathbf{a}+\mathbf{b})+\mathbf{c} \quad \ldots(2)$$

From triangle PQR, we have

$$\vec{PR}=\vec{PQ}+\vec{QR}=\mathbf{b}+\mathbf{c} \quad \ldots(3)$$

From triangle OPR, we have

$$\vec{OR}=\vec{OP}+\vec{PR}=\mathbf{a}+(\mathbf{b}+\mathbf{c}) \quad \ldots(4)$$

From (4) and (2) we have, by equating the two expressions for $\vec{OR}$,

$$\boxed{\mathbf{a}+(\mathbf{b}+\mathbf{c})=(\mathbf{a}+\mathbf{b})+\mathbf{c}}$$

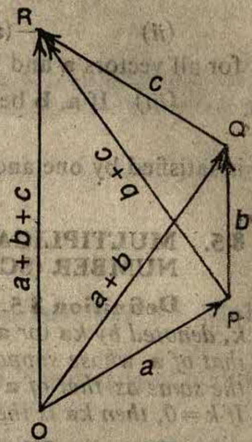


Fig. 8·7.

Remarks 1. In view of theorems 8·1 and 8·2, the sum of any number of vectors is independent of their order and grouping. In particular, the vectors $\mathbf{a}+(\mathbf{b}+\mathbf{c})$, $\mathbf{a}+(\mathbf{c}+\mathbf{b})$, $\mathbf{b}+(\mathbf{c}+\mathbf{a})$, $\mathbf{b}+(\mathbf{a}+\mathbf{c})$, $\mathbf{c}+(\mathbf{a}+\mathbf{b})$, $\mathbf{c}+(\mathbf{b}+\mathbf{a})$, $(\mathbf{a}+\mathbf{b})+\mathbf{c}$, $(\mathbf{b}+\mathbf{a})+\mathbf{c}$, $(\mathbf{c}+\mathbf{a})+\mathbf{b}$, $(\mathbf{a}+\mathbf{c})+\mathbf{b}$, $(\mathbf{b}+\mathbf{c})+\mathbf{a}$, $(\mathbf{c}+\mathbf{b})+\mathbf{a}$ are all equal, and we can denonte all of them by a common symbol, namely, $\mathbf{a}+\mathbf{b}+\mathbf{c}$.

2. If **b** be any vector and **0** be the zero vector, then

$$\boxed{\mathbf{a}+\mathbf{0}=\mathbf{0}+\mathbf{a}} \quad \ldots(A)$$

8·4 1. Negative of a Vector

Definition 8·4. *The vector which has the same mangitude and same support as* **a**, *but the opposite sense, is called the negative of* **a** *and is denoted by* $-\mathbf{a}$. *It can be easily seen that*

$$\boxed{\mathbf{a}+(-\mathbf{a})=(-\mathbf{a})+\mathbf{a}=\mathbf{0}} \quad \ldots(B)$$

8·4·2. Subtraction of Vectors

Let **a** and **b** be two vectors. The vector $\mathbf{a}+(-\mathbf{b})$ is usually denoted by $\mathbf{a}-\mathbf{b}$. We thus get another operation on vectors, namely subtraction. It can be easily seen that :

(i) $\mathbf{a}-\mathbf{a}=\mathbf{0}$, $\mathbf{0}-\mathbf{a}=-\mathbf{a}$,

for all vectors **a**, and

(*ii*) $\qquad -(\mathbf{a}+\mathbf{b})=(-\mathbf{a})+(-\mathbf{b}),$

for all vectors **a** and **b**.

(*iii*) If **a**, **b** be two given vectors, then the vector equation

$$\mathbf{x}+\mathbf{b}=\mathbf{a}$$

is satisfied by one and only one vector

$$\mathbf{x}=\mathbf{a}-\mathbf{b}.$$

8·5. MULTIPLICATION OF A VECTOR BY A REAL NUMBER (SCALAR)

Definition 8·5. *The product of a vector* **a** *and a real number* k*, denoted by* $k\mathbf{a}$ (*or* $\mathbf{a}k$) *is a vector whose magnitude is* $|k|$ *times that of* **a**, *whose support is the same as that of* **a**, *and whose sense is the same as that of* **a** *or the opposite one according as* $k>0$ *or* $k<0$. *If* $k=0$, *then* $k\mathbf{a}$ *is the zero vector.*

Theorem 8·3. *If* **a** *be any vector, and* k, l *be any real numbers, then*

(*i*) $\qquad (-1)\mathbf{a}=-\mathbf{a}\,;$

(*ii*) $\qquad k(-\mathbf{a})=(-k)\mathbf{a}=-(k\mathbf{a})\,;$

(*iii*) $\qquad (-k)(-\mathbf{a})=k\mathbf{a}\,;$

(*iv*) $\qquad k(l\mathbf{a})=(kl)\mathbf{a}.$

Proof. (*i*) The magnitude of $(-1)\,\mathbf{a}$ is $|-1|\,|\mathbf{a}|$, *i.e.*, the same as that of **a**, and the direction of $(-1)\,\mathbf{a}$ is opposite to that of **a**. Hence $(-1)\,\mathbf{a}=-\mathbf{a}$.

(*iii*) If $k>0$, the magnitude of

$$(-k)(-\mathbf{a})=|-k|\,|-\mathbf{a}|=k\,|\mathbf{a}|=|k\mathbf{a}|,$$

and direction of $(-k)(-\mathbf{a})$ is opposite to that of $-\mathbf{a}$, *i.e.*, the same as that of **a**, which is also the direction of $k\mathbf{a}$.

Hence $\qquad (-k)(-\mathbf{a})=k\mathbf{a}.$

If $k<0$, then $-k>0$, and so magnitude of

$$\begin{aligned}(-k)(-\mathbf{a})&=(-k)\,|-\mathbf{a}|,\\ &=(-k)\,|\mathbf{a}|,\\ &=|-k\mathbf{a}|,\\ &=|k\mathbf{a}|,\end{aligned}$$

and direction of $(-k)(-\mathbf{a})$ is the same as that of $-\mathbf{a}$, *i.e.*, opposite of that of **a**, *i.e.*, the same as that of $k\mathbf{a}$.

Hence, in this case also

$$(-k)(-\mathbf{a})=k\mathbf{a}.$$

If $k=0$, the vectors $(-k)(-\mathbf{a})$ and $k\mathbf{a}$ are both **0**, and so are equal.

(*ii*) and (*iv*). Similar to that of (*iii*).

The relation between vector addition and multiplication of a vector by a real number is given by the following theorem which we state without proof.

Theorem 8·4. *Let* **a**, **b** *be vectors and k, l be real numbers. Then*

(*i*) $k(\mathbf{a}+\mathbf{b})=k\mathbf{a}+k\mathbf{b}$;

(*ii*) $(k+l)\mathbf{a}=k\mathbf{a}+l\mathbf{a}$.

8·5·1. Basic Properties of Addition of Vectors and of Multiplication of a Vector by a Scalar

We list below, for ready reference, the basic laws of addition of vectors and of multiplication of vectors by scalars.

Let us denote the set of real numbers by **R** (as usual) and the set of all vectors by **V**. Then,

I $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$ for all $\mathbf{a}, \mathbf{b}\in\mathbf{V}$.

II $\mathbf{a}+(\mathbf{b}+\mathbf{c})=(\mathbf{a}+\mathbf{b})+\mathbf{c}$, for all $\mathbf{a}, \mathbf{b}, \mathbf{c}\in\mathbf{V}$.

III There exists a vector, namely **0**, such that

$\mathbf{a}+\mathbf{0}=\mathbf{0}+\mathbf{a}=\mathbf{a}$, for all $\mathbf{a}\in\mathbf{V}$,

IV To each $\mathbf{a}\in\mathbf{V}$ there corresponds $\mathbf{b}\in\mathbf{V}$ such that

$$\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}=\mathbf{0}.$$

V $(mn)\mathbf{a}=m(n\mathbf{a})$ for all $m, n\in\mathbf{R}$,

VI $1(\mathbf{a})=\mathbf{a}$, for all $\mathbf{a}\in\mathbf{V}$.

VII $m(\mathbf{a}+\mathbf{b})=m\mathbf{a}+m\mathbf{b}$, for all $m\in\mathbf{R}$ and $\mathbf{a}, \mathbf{b}\in\mathbf{V}$.

VIII $(m+n)\mathbf{a}=m\mathbf{a}+n\mathbf{a}$ for all $m, n\in\mathbf{R}$, and $\mathbf{a}\in\mathbf{V}$.

8·6. LINEAR COMBINATION OF VECTORS

A vector **r** is said to be a linear combination of vectors **a**, **b**, **c** etc., if there exist scalars x, y, z etc. such that

$$\mathbf{r}=x\mathbf{a}+y\mathbf{b}+z\mathbf{c}$$

For example, the vectors $2\mathbf{a}-3\mathbf{b}+\mathbf{c}$ is a linear combination of the vectors **a**, **b** and **c**.

If two vectors have the same or parallel supports, then each is a linear combination of the other. In other words, if **a**, **b** be two parallel vectors, then there exists a scalar t such that $\mathbf{b}=t\mathbf{a}$, or $\mathbf{a}=(1/t)\mathbf{b}$.

8·6·1. Coplanar Vectors

A set of vectors is said to be *coplanar*, if their supports are parallel to the same plane, *i.e.*, if there exists a plane parallel to the supports of all of them. The supports of a set of coplanar co-initial vectors are coplanar.

8·6·2. Vectors in Two Dimensions

Let **a**, **b** be two given non-collinear vectors. Every vector **r** coplanar with **a** and **b** can be uniquely expressed as a linear combination $x\mathbf{a}+y\mathbf{b}$, for some scalars x and y.

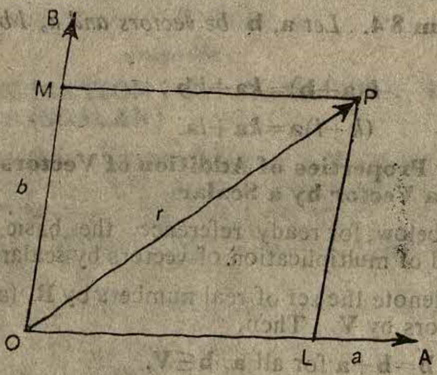


Fig. 8·8.

Existence. Without loss of generality we can assume that the vectors **a**, **b** and **r** are coinitial. Take a point O. Let the vectors **a**, **b**, **r** be $\overrightarrow{OA}$, $\overrightarrow{OB}$ and $\overrightarrow{OP}$ respectively. Since the vectors **a**, **b**, **r** are coplanar, therefore the lines OA, OB, OP are coplanar.

Through P draw PL ∥ BO to meet OA in L, and PM ∥ AO to meet OB in M. Now

$$\mathbf{r}=\overrightarrow{OP}=\overrightarrow{OL}+\overrightarrow{LP}=\overrightarrow{OL}+\overrightarrow{OM} \quad \ldots(1)$$

Since $\overrightarrow{OL}$ and $\overrightarrow{OA}$ are collinear vectors, therefore there exists a scalar x such that $\overrightarrow{OL}=x\mathbf{a}$. Similarly there exists a scalar y such that $\overrightarrow{OM}=y\mathbf{b}$. We can, therefore write (1) as

$$\mathbf{r}=x\mathbf{a}+y\mathbf{b}.$$

Uniqueness. Let, if possible, $\mathbf{r}=x\mathbf{a}+y\mathbf{b}$, and also $\mathbf{r}=x'\mathbf{a}+y'\mathbf{b}$. Then $x\mathbf{b}+y\mathbf{b}=x'\mathbf{a}+y'\mathbf{b}$, so that

$$x\mathbf{a}+y\mathbf{b}=x'\mathbf{a}+y'\mathbf{b}$$

or $\quad (x-x')\mathbf{a}+(y-y')\mathbf{b}=0.$

If $x-x'\neq 0$, we have

$$\mathbf{a}=-\frac{y-y'}{x-x'}\mathbf{b},$$

so that **a** and **b** are collinear. Since **a** and **b** are given to be non-collinear, we have a contradiction. Therefore $x=x'$.

Similarly $y=y'$. Hence the representation is unique.

Remarks 1. In the above discussion $x\mathbf{a}$ is called the component of **r** in the direction of **a** along $y\mathbf{b}$ is called the component of **r** in the direction of **b**.

2. If **a** and **b** are perpendicular to each other, then we say that $x\mathbf{a}$ is called the resolved part of **r** and **a** along $y\mathbf{b}$ is called the resolved part of **r** along **b**.

3. If **a** and **b** are unit vectors perpendicular to each other, then the co-ordinates of P referred to OA and OB as the axes of co-ordinates are (x, y). For this reason, we often say that (x, y) are the co-ordinates of the vector relative to the vectors **a** and **b** even when **a** and **b** are not orthogonal vectors.

Vectors in three dimensions. Components of a vector with respect to three non-coplanar vectors.

8·6·3. Components of a Vector

Any vector **r** can be expressed as the sum of three vectors, parallel to *any* three non-coplanar vectors. Let **a**, **b**, **c** be three non-coplanar vectors. With any point O as origin of reference, take

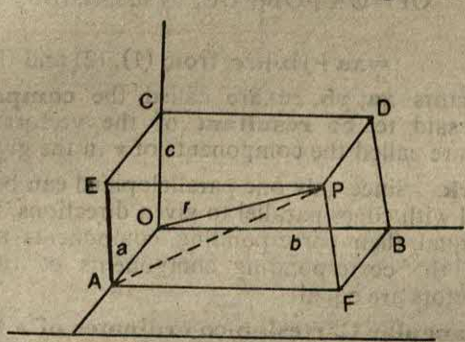


Fig. 8·9.

$\overrightarrow{OP}=\mathbf{r}$, and on OP as a diagonal construct a parallelopiped with edges OA, OB, OC parallel to **a**, **b**, **c** respectively.

Since $\overrightarrow{OA}$ is parallel to **a**, there exists a real number x such that

$$\overrightarrow{OA}=x\mathbf{a}. \qquad \text{--(1)}$$

Since $\overrightarrow{OB}$ is parallel to **b**, there exists a real number y such that

$$\overrightarrow{OB}=y\mathbf{b}. \quad ...(2)$$

Again, since $\overrightarrow{OC}$ is parallel to $\mathbf{c}$, there exists a real number z such that

$$\overrightarrow{OC}=z\mathbf{c}. \quad ...(3)$$

From $\triangle OAP$, we have

$$\overrightarrow{OP}=\overrightarrow{OA}+\overrightarrow{AP},$$

$$=\overrightarrow{OA}+(\overrightarrow{AF}+\overrightarrow{FP}), \text{ from } \triangle AFP,$$

$$=\overrightarrow{OA}+(\overrightarrow{OB}+\overrightarrow{OC}),$$

since $\overrightarrow{AF}=\overrightarrow{OB}$ (AF, OB being opposite sides of the parallelogram AFBO), and $\overrightarrow{FP}=\overrightarrow{AE}=\overrightarrow{OC}$ (why ?).

$\therefore \quad \overrightarrow{OP}=\overrightarrow{OA}+\overrightarrow{OB}+\overrightarrow{OC}$, by *associativity* of addition of vectors,

$$=x\mathbf{a}+y\mathbf{b}+z\mathbf{c}, \text{ from (1), (2) and (3).}$$

The vectors $x\mathbf{a}$, $y\mathbf{b}$, $z\mathbf{c}$ are called the **component vectors** of $\mathbf{r}$ and $\mathbf{r}$ is said to be **resultant** of the vectors $x\mathbf{a}$, $y\mathbf{b}$ and $z\mathbf{c}$. Also, x, y, z are called the components of $\mathbf{r}$ in the given directions.

Remark. Since only one parallelopiped can be drawn on OP as a diagonal with edges parallel to given directions, therefore if two vectors are equal, their corresponding components are equal, and *conversely* if the corresponding components of two vectors are equal, the vectors are equal.

8 6·4. Rectangular Cartesian co-ordinates of a Point

We know that the position of a point in a plane can be specified by choosing a point on the plane as the origin and a pair of perpendicular lines through O as the axes of co-ordinates. In a similar way, the position of a point in space can be specified by choosing a point in space as the origin and three mutually perpendicular lines as the axes of co-ordinates.

Given a line, by choosing any arbitrary point on the line as the origin, a suitable distance as a unit, and one of the two directions on the line as the positive direction, we can set up a one-to-one correspondence between the set of real numbers and the set of points on the line. In other words, we can set up a co-ordinate system for points on the line, each point having one co-ordinate.

Take any point O and three mutually perpendicular straight lines X′OX, Y′OY, Z′OZ through O.

Let us set up a co-ordinate system for points on the lines X′OX, Y′OY, Z′OZ in such a manner that O is the origin of reference for each line, the unit is the same for all the three lines and the directions OX, OY, OZ are taken as the positive directions along the lines X′OX, Y′OY, Z′OZ respectively. In other words, we set up the convention that the co-ordinates of all points on OX, OY and OZ (except O) will be positive, and the co-ordinates of all points on OX′, OY′, OZ′ (except O) will be negative.

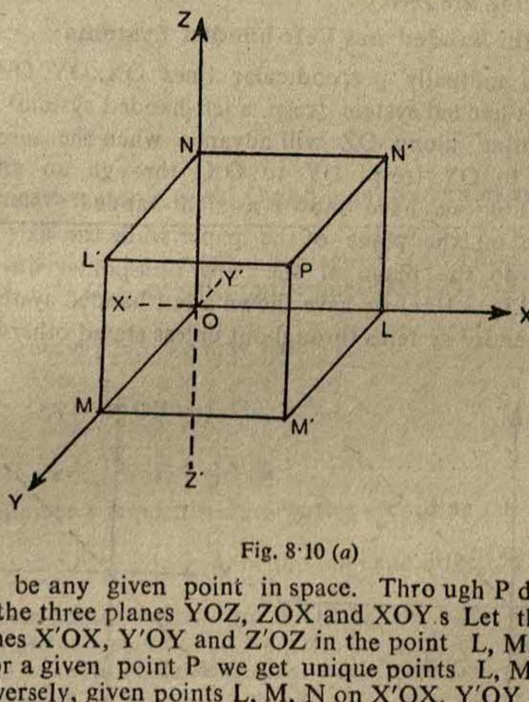


Fig. 8·10 (*a*)

Let P be any given point in space. Thro ugh P draw planes parallel to the three planes YOZ, ZOX and XOY s Let these planes meet the lines X′OX, Y′OY and Z′OZ in the point L, M, N respectively. For a given point P we get unique points L, M, N in this way. Conversely, given points L, M, N on X′OX, Y′OY and Z′OZ respectively, we can determine a unique point P in space by drawing planes through the points L, M and N parallel to the planes YOZ, ZOX and XOY respectively. Using the co-ordinate system (set up above) on the lines X′OX, Y′OY, Z′OZ, we can associate unique real numbers x, y and z with the points L, M and N respectively. In this manner, to each point P there corresponds exactly one ordered triple of real numbers (x, y, z) and to every such ordered triple there corresponds exactly one point P. We thus get a one-to-one correspondance between the points of space an ordered triples of real numbers. We call this correspondence a *rectangular cart sian co-ordinate system.* The lines X′OX, Y′OY, Z′OZ are called the axis of x, the axis of y and the axis of z respectively.

The planes YOZ, ZOX and XOY are called the yz-plane, zx-plane and xy-plane respectively. Also, (x, y, z) are called the cartesian co-ordinates of the point P.

The x-co-ordinate of every point on the yz-plane is zero, the y-co-ordinate of every point on the zx-plane is zero, and the z-co-ordinate of every point on the xy-plane is zero. The y and z co-ordinates of every point on X′OX are zero ; the z and x-co-ordinates of every point on Y′OY are zero ; the x and y co-ordinates of every point on Z′OZ are zero.

8·6·5. Right-handed and Left-handed Systems

Three mutually perpendicular lines OX, OY, OZ are said to form a right-handed system (resp. a left-handed system) provided a screw pointing along OZ will advance when the screw is twisted from OX to OY (resp. OY to OX) through an angle of 90°. In fig. 8·10 (*b*) we have shown a right-handed system. The lines OY, OZ lie on the plane of the paper while the axis of X is perpendicular to the plane of the paper and points out towards the reader. In Fig. 8·10(*c*) we have shown a left-handed system. We shall use right-handed systems throughout unless stated otherwise.

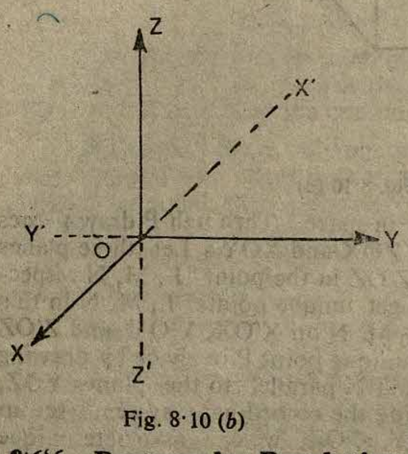


Fig. 8·10 (*b*)

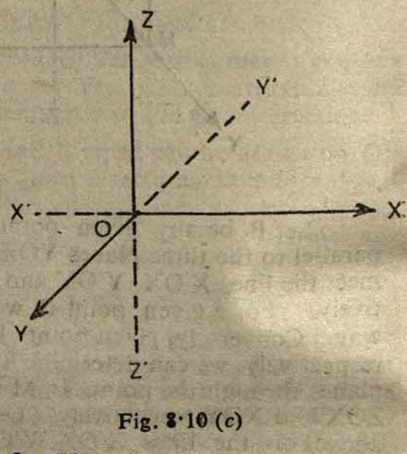


Fig. 8·10 (*c*)

8·6·6. Rectangular Resolution of a Vector

A special case, and the most useful one, of resolution of a vector into component vectors is that in which the three directions are mutually at right angles.

Consider a right-handed system of mutually perpendicular axes OX, OY and OZ (perpendicular to each other). Let the unit vectors parallel to OX, OY and OZ be denoted by **i**, **j** and **k** respectively.

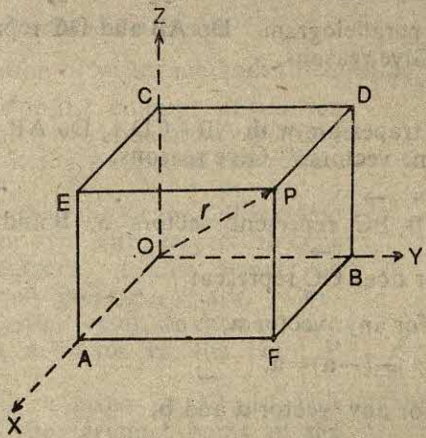


Fig. 8·10 (*d*)

Consider a vector **r** with O and P as the initial and terminal points respectively.

The parallelopiped on OP as a diagonal, and edges parallel to the vectors **i**, **j**, **k** (*i.e.*, parallel to the lines OX, OY, OZ) is a rectangular one.

As in section 8·6·3, we may write

$$\overrightarrow{OA}=x\mathbf{i},\ \overrightarrow{OB}=y\mathbf{j},\ \overrightarrow{OC}=z\mathbf{k},$$

$$\mathbf{r}=\overrightarrow{OP}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}.$$

The number x is positive or negative according as the vector $\overrightarrow{OA}$ has the same direction as **i**, or the opposite direction. Similarly y is positive or negative according as $\overrightarrow{OB}$ has the same direction as **j**, or the opposite direction, and similarly for z. The numbers x, y, z are called the **rectangular components, resolutes, or resolved parts** of **r** for the directions **i**, **j** and **k**.

The resolution of a vector **r** in terms of **i**, **j** and **k** is unique.

EXERCISE 8 (*a*)

1. ABCD is a square. Do $\overrightarrow{AB}$ and $\overrightarrow{CD}$ represent the same vectors ? Give reasons.

2. ABCD is a rectangle. Do $\overrightarrow{AD}$ and $\overrightarrow{BC}$ represent the same vectors ? Give reasons.

3. ABCD is a parallelogram. Do $\overrightarrow{AB}$ and $\overrightarrow{DC}$ represent the same vectors ? Give reasons.

4. ABCD is a trapezium with AB ∥ CD. Do $\overrightarrow{AB}$ and $\overrightarrow{CD}$ represent the same vectors ? Give reasons.

5. Let $\overrightarrow{OA}$, $\overrightarrow{AB}$, $\overrightarrow{BC}$ represent vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ respectively. What vector does $\overrightarrow{OC}$ represent ?

6. Prove that for any vector $\mathbf{a}$,

$$-(-\mathbf{a})=\mathbf{a}.$$

7. Show that for any vector $\mathbf{a}$ and $\mathbf{b}$,

$$-(\mathbf{a}+\mathbf{b})=(-\mathbf{a})+(-\mathbf{b}).$$

8. ABCD is a square. If $\overrightarrow{AB}=\mathbf{a}$, $\overrightarrow{BC}=\mathbf{b}$, show that

(*i*) $\overrightarrow{AC}=\mathbf{a}+\mathbf{b}$, (*ii*) $\overrightarrow{BD}=\mathbf{b}-\mathbf{a}$,

(*iii*) $\overrightarrow{DA}=-\mathbf{b}$.

9. G is the mid-point of a triangle ABC. Show that

$$\overrightarrow{GA}+\overrightarrow{GB}+\overrightarrow{GC}=\mathbf{0}.$$

10. D, E, F are the mid-points of the sides BC, CA, AB respectively of a triangle ABC. Prove that

$$\overrightarrow{AD}+\overrightarrow{BE}+\overrightarrow{CF}=\mathbf{0}.$$

11. ABCDEF is a regular hexagon. If $\overrightarrow{AB}=\mathbf{a}$, $\overrightarrow{BC}=\mathbf{b}$, express $\overrightarrow{AC}$, $\overrightarrow{CD}$, $\overrightarrow{BE}$ in terms of $\mathbf{a}$ and $\mathbf{b}$.

12. ABCDEF is a regular hexagon. Prove that

$$\overrightarrow{AB}+\overrightarrow{AC}+\overrightarrow{AD}+\overrightarrow{AE}+\overrightarrow{AF}=3\overrightarrow{AD}.$$

13. E, F are the middle points of the diagonals AC, BD respecpectively of a quadrilateral ABCD. Prove that

$$\overrightarrow{AB}+\overrightarrow{AD}+\overrightarrow{CB}+\overrightarrow{CD}=4\overrightarrow{EF}.$$

8·7. APPLICATION OF VECTORS TO GEOMETRY

8·7·1. Position Vector of a Point

The application of vector methods to solution of geometrical problems depends on the concept of position vector of a point. We choose an arbitrary point O and call it the **origin of reference**. If P be any point, then the vector $\overrightarrow{OP}$ is called the position vector of P with respect to O as the origin of reference. In this manner we can associat to every point P a vector. Conversely, to every given vector, we can associate a point. Given a vector $\mathbf{r}$, the point P associated with it is given by $\overrightarrow{OP}=\mathbf{r}$. This means that P is the terminal point of the vector $\mathbf{r}$ having O as its initial point. Thus, by choosing an origin of reference O we can set up a one-to-one correspondence between the set of all vectors having O as the initial point, and the set of all points in space.

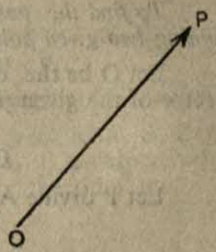


Fig. 8·11.

For different choices of the origin of reference, the same point will have different position vectors.

The origin of reference can be chosen arbitrarily, but by a suitable choice of the origin of reference, the solution of a geometrical problem is often very much simplified.

8·7·2. Expression for a Vector in Terms of Position Vectors

By using the concept of position vector of a point, it is possible to express a vector in terms of the position vectors of the end points.

Let $\overrightarrow{AB}$ be a vector and let O be the origin of reference. If the position vectors of A and B be $\mathbf{a}$ and $\mathbf{b}$ respectively, then from the equality

$$\overrightarrow{AB}=\overrightarrow{OB}-\overrightarrow{OA}$$

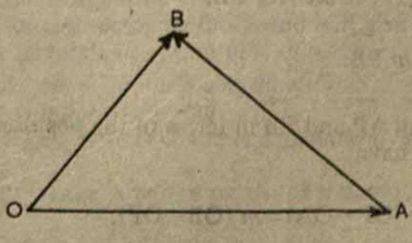


Fig. 8·12.

we find that $\vec{AB}=\mathbf{b}-\mathbf{a}$. Such an expression is found to be very useful in geometry.

8·7·3. Section Formula

To find the position vector of the point which divides the line joining two given points in a given ratio.

Let O be the origin of reference and let **a, b** be the position vector of the given points A, B so that we have

$$\vec{OA}=\mathbf{a},\ \vec{OB}=\mathbf{b}.$$

Let P divide AB so that

$$\frac{AP}{PB}=\frac{m}{n} \qquad ...(1)$$

The ratio m/n is positive if P divides AB internally, and negative if P divides AB externally. We shall express the position vector $\vec{OP}$ of the point P in terms of the position vectors $\vec{OA}$ and $\vec{OB}$ of the points A and B.

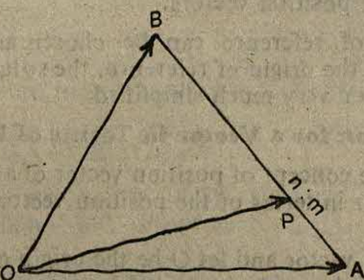


Fig. 8·13.

We can write (1) as

$$n\ AP=m\ PB,$$

so that $n\ \vec{AP}=m\ \vec{PB}$.

Expressing $\vec{AP}$ and $\vec{PB}$ in terms of the position vectors of the end points, we have

$$n\ (\vec{OP}-\vec{OA})=m\ (\vec{OB}-\vec{OP}),$$

or $$(n+m)\ \vec{OP}=m\ \vec{OB}+n\ \vec{OA},$$

or
$$\vec{OP} = \frac{m\,\vec{OB}+n\vec{OA}}{m+n},$$
$$=\frac{m\,\mathbf{b}+n\,\mathbf{a}}{m+n}.$$

Corollary. The position vector of the mid-point of the join of two points with position vectors **a** and **b** is $\frac{1}{2}(\mathbf{a}+\mathbf{b})$.

Example 1. *Show that the medians of a triangle are concurrent.*

Solution. Let the position vectors of the vertices A, B, C of a triangle ABC with respect to any origin O be **a**, **b**, **c** respectively. The position vectors of the mid-points D, E, F of the sides are

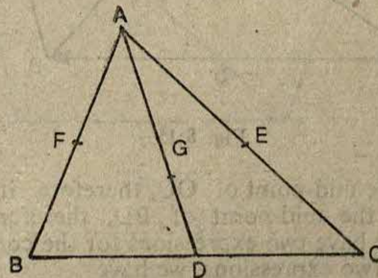


Fig. 8·14.

$\frac{1}{2}(\mathbf{b}+\mathbf{c})$, $\frac{1}{2}(\mathbf{c}+\mathbf{a})$ and $\frac{1}{2}(\mathbf{a}+\mathbf{b})$ respectively.

The point G dividing AD in the ratio 2 : 1 is

$$\frac{2.\,(\mathbf{b}+\mathbf{c})+1.\,\mathbf{a}}{2+1}=\frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c}).$$

By symmetry, this point also lies on the other two medians.

Hence the medians of a triangle are concurrent,

Remark. In class XI we had proved this result using the method of co ordinate geometry. Now we have a co-ordinate free proof of the same result. This is not surprising. We shall prove several known results again by vector methods. This will enable us to see the power of the vector methods and will give us practice in the use of vector methods to obtain new results.

Example 2. *Show that a plane quadrilateral whose diagonals bisect each other is a parallelogram.*

Solution. Let ABCD be a quadrilateral whose diagonals bisect each other. Take A as the origin of reference. Let $\vec{AB}=\mathbf{b}$, and $\vec{AD}=\mathbf{d}$, so that **b**, **c** and **d** are the position vectors of the points B, C and D respectively. Since the diagonals bisect each other,

therefore O, the point of intersection of the diagonals is the mid-point of AC as well as BD.

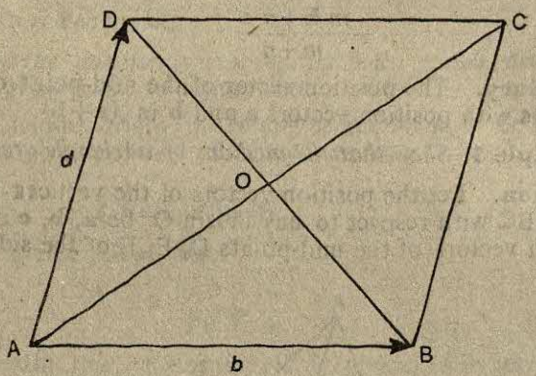


Fig. 8·15.

Since O is the mid-point of OC, therefore it is the point $\frac{1}{2}$ **c**. Again, since O is the mid-point of BD, therefore it is the point $\frac{1}{2}(\mathbf{b}+\mathbf{d})$. We now have two expressions for the common mid-point O. Equating the two expressions, we have

$$\frac{1}{2}\,\mathbf{c}=\frac{1}{2}(\mathbf{b}+\mathbf{d}),$$

or $$\mathbf{c}=\mathbf{b}+\mathbf{d}$$

so that $$\mathbf{d}=\mathbf{c}-\mathbf{b},$$

i.e., $$\overrightarrow{AD}=\overrightarrow{AC}-\overrightarrow{AB},$$

i.e., $$\overrightarrow{AD}=\overrightarrow{BC}.$$

This means that AD is equal as well as parallel to BC, and consequently ABCD is a parallelogram.

Remark. Try to prove the above result by choosing O as the origin of reference, and also by choosing an arbitrary origin of reference.

Example 3. *Show that the straight line joining the mid-points of the diagonals of a trapezium is parallel to the parallel sides and is half of their difference.*

Solution. Let ABCD be a trapezium with parallel sides AB, CD. Take A as the origin of reference, and let the position vectors of B and D be **b** and **d** respectively. Since DC is parallel to AB,

therefore the vector $\overrightarrow{DC}$ must be a product of the vector $\overrightarrow{AB}$ by some scalar, say t. Let $\overrightarrow{DC}=t\,\overrightarrow{AB}=t\,\mathbf{b}$.

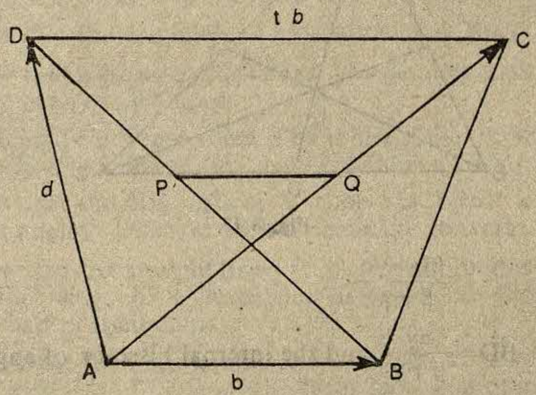


Fig. 8·16.

$\therefore$ The position vector of C is

$$\overrightarrow{AC}=\overrightarrow{AD}+\overrightarrow{DC}=\mathbf{d}+t\,\mathbf{b}.$$

The position vector of the mid-points P, Q of the diagonals BD and AC are $\frac{1}{2}$ $(\mathbf{b}+\mathbf{d})$ and $\frac{1}{2}(\mathbf{d}+t\,\mathbf{b})$ respectively.

Now
$$\overrightarrow{PQ}=\overrightarrow{AQ}-\overrightarrow{AP}$$
$$=\frac{1}{2}\,(\mathbf{d}+t\,\mathbf{b})-\frac{1}{2}\,(\mathbf{b}+\mathbf{d})$$
$$=\frac{1}{2}\,(t-1)\,\mathbf{b}=\frac{1}{2}\,(t-1)\,\overrightarrow{AB}$$

Thus PQ is parallel to AB. Also the length of PQ is $\frac{1}{2}\mid t-1\mid$ AB *i.e.*, half the difference of the lengths of AB and CD. (Observe that the length of CD is simply t. AB, so that the difference of the two lengths is $\mid t-1\mid$ AB).

Example 4. *Show that the internal bisectors of the angles of a triangle are concurrent.*

Solution. Take any point O as the origin of reference. Let **a**, **b**, **c** be the position vectors of A, B and C, and let α, β, γ be the lengths of the sides of the triangle. The position vector of the point D where the internal bisector of the angle A meets BC is

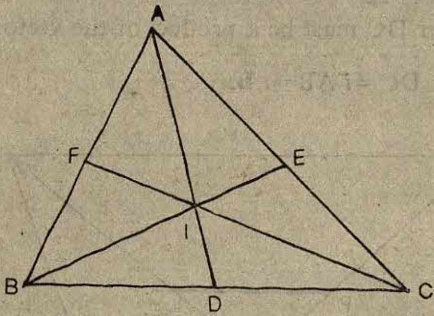


Fig. 8·17.

$$\frac{\beta\mathbf{b}+\gamma\mathbf{c}}{\beta+\gamma}.$$

Now $\quad BD=\dfrac{\alpha\gamma}{\beta+\gamma}$ and the internal bisector of angle B will divide DA in the ratio BD : BA, *i.e.*, $\dfrac{\alpha\gamma}{\beta+\gamma}:\gamma$, *i.e.*, $\alpha:\beta+\gamma$. The position vector of the point I dividing DA in the ratio $\alpha:\beta+\gamma$ is

$$\frac{\alpha\mathbf{a}+\beta\mathbf{b}+\gamma\mathbf{c}}{\alpha+\beta+\gamma}. \qquad \ldots(1)$$

By symmetry this point lies on the other two bisectors as well. Thus the bisectors are concurrent, and the position vector of the point of concurrence is given by (1).

Example 5. *Show that the straight lines joining the mid-points of pairs of opposite edges of a tetrahedron are concurrent.*

Solution. Let the position vectors of the vertices of a tetrahedron ABCD, with respect to any origin of reference O, be **a**, **b**, **c**, **d** respectively. The position vectors of the mid-points P and Q, of BD and AC respectively are $\frac{1}{2}$ $(\mathbf{a}+\mathbf{d})$ and $\frac{1}{2}$ $(\mathbf{a}+\mathbf{c})$. The position vector of G, the mid-point of PQ, is

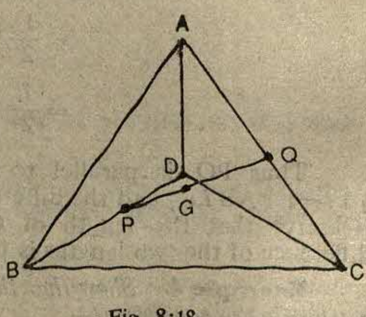


Fig. 8·18.

$$\frac{1}{2}\left[\frac{1}{2}(\mathbf{b}+\mathbf{d})+\frac{1}{2}(\mathbf{a}+\mathbf{c})\right]=\frac{1}{4}(\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d}).$$

By symmetry we find that this point also lies on the straight lines joining the mid-points of other pairs of opposite sides.

EXERCISE 8 (*b*)

1. Show that the line joining the mid-points of two sides of a triangle is parallel to the third side and half of it.
2. Show that the diagonals of a parallelogram bisect each other.
3. Show that a plane quadrilateral whose diagonals bisect each other is a parallelogram.
4. Show that the figure formed by joining the middle points of the adjacent sides of any quadrilateral is a parallelogram.
5. Show that the diagonals of a rhombus bisect each other at right angles. Also, state and prove the converse of this result.
6. Show that the straight line joining the mid-points of two non-parallel sides of a trapezium is parallel to the parallel sides and half of their sum.
7. Show that the lines joining the vertices of a tetrahedron to the centroids of the opposite faces meet in a point.
8. Show that the four diagonals of any parallelopiped are concurrent and are bisected at the point of concurrence.

8·8. SCALAR PRODUCT OF TWO VECTORS

Definition 8·6. *The* **scalar product** *of two vectors* **a** *and* **b**, *to be denoted by* **a** . **b** *is the scalar*

$$\mathbf{a}\,.\,\mathbf{b} \equiv |\,\mathbf{a}\,|\;\;|\,\mathbf{b}\,|\,\cos\theta,$$

where θ *is the angle between* **a** *and* **b**.

Since we use a dot (.) to indicate the scalar product of vectors, therefore scalar product is often called **dot product**.

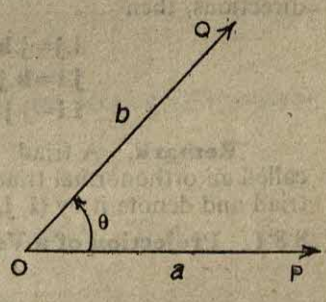


Fig. 8·19.

Illustrations. (*i*) Let **a** and **b** be two unit vectors inclined to each other at an angle of 60°. Then

$$\mathbf{a}\,.\,\mathbf{b} = |\,\mathbf{a}\,|\,.\,|\,\mathbf{b}\,|\cos 60°,$$

$$=1\,.\,1\,.\frac{1}{2}=\frac{1}{2}, \text{ since } |\,\mathbf{a}\,| = |\,\mathbf{b}\,| = 1.$$

(*ii*) Let **a** and **b** be two vectors such that

$$|\,\mathbf{a}\,| = 2,\; |\,\mathbf{b}\,| = 3,$$

and the angle between them is 45°. Then

$$\mathbf{a}\,.\,\mathbf{b} = |\,\mathbf{a}\,|\,.\,|\,\mathbf{b}\,|\cos 45°,$$

$$=2\,.\,3\,.\frac{1}{\sqrt{2}}=3\sqrt{2}.$$

The following particular cases of scalar product are interesting:

(*i*) If **a** and **b** are in the same direction, then $\theta=0$, and

$$\mathbf{a}\cdot\mathbf{b}=|\mathbf{a}|\cdot|\mathbf{b}|.$$

(*ii*) If **a** and **b** are parallel but in opposite directions, then $\theta=\pi$, and

$$\mathbf{a}\cdot\mathbf{b}=-|\mathbf{a}|\cdot|\mathbf{b}|,$$

(*iii*) If **a** and **b** are perpendicular, then $\theta=\frac{\pi}{2}$, and

$$\mathbf{a}\cdot\mathbf{b}=0.$$

Thus **a** . **b** can be zero even when **a** and **b** are both different from zero.

(*iv*) If $\mathbf{a}=\mathbf{b}$, then $|\mathbf{a}|=|\mathbf{b}|$, and $\theta=0$, so that

$$\mathbf{a}\cdot\mathbf{a}=(|\mathbf{a}|)^2,$$

and consequently $|\mathbf{a}|=\sqrt{(\mathbf{a}\cdot\mathbf{a})}$.

We thus have an expression for $|\mathbf{a}|$ in terms of scalar product of **a** with itself.

Remark. It is usual to denote **a** . **a** by $\mathbf{a}^2$. We shall adopt this notation.

(*v*) If **i**, **j**, **k** are unit vectors in three mutually perpendicular directions, then

$$\mathbf{i}.\mathbf{j}=\mathbf{j}.\mathbf{k}=\mathbf{k}.\mathbf{i}=0,$$
$$\mathbf{j}.\mathbf{i}=\mathbf{k}.\mathbf{j}=\mathbf{i}.\mathbf{k}=0,$$
$$\mathbf{i}.\mathbf{i}=\mathbf{j}.\mathbf{j}=\mathbf{k}.\mathbf{k}=1.$$

Remark. A triad of mutually orthogonal unit vectors is called an orthonormal triad. We shall often use an orthonormal triad and denote it by (**i**, **j**, **k**).

8·8·1. Projection of a Vector on a Directed Line

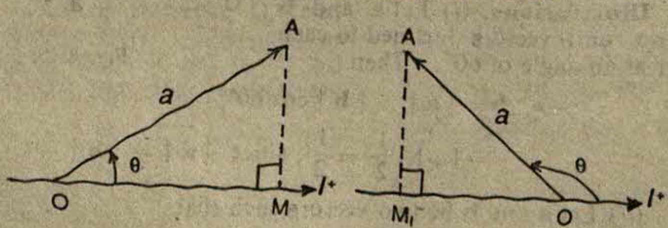


Fig. 8·20.

Definition 8·7. *Let l be a directed line, and* **a**, *a given vector. Let O be any point on l and let* $\vec{OA}$ *represent* **a**. *Let M be the foot of the perpendicular from A on l.* $|\mathbf{a}|\cos\theta$ *is called the* **projection** *of* **a** *on l, where* θ *is a angle of* $l^{+}OA$.

Remark. The vector $\overrightarrow{OM}$ in the above discussion is sometimes called the **vector projection** of **a** on l. To avoid confusion in such a situation, projection as defined above is called **scalar projection.**

If **a** and **b** be two vectors inclined to each other at an angle θ, then

projection of **a** on $\mathbf{b} = |\mathbf{a}| \cos\theta$,

projection of **b** on $\mathbf{a} = |b| \cos\theta$,

$$\begin{aligned}\mathbf{a} \cdot \mathbf{b} &= |\mathbf{a}|\ |\mathbf{b}| \cos\theta,\\ &= |\mathbf{a}|\ (\ |\mathbf{b}| \cos\theta),\\ &= |\mathbf{a}|\ (\text{projection of } \mathbf{b} \text{ on } \mathbf{a}),\\ &= |\mathbf{b}|\ (\ |\mathbf{a}| \cos\theta),\\ &= |\mathbf{b}|\ (\text{projection of } \mathbf{a} \text{ on } \mathbf{b}).\end{aligned}$$

Corollary. For any two vectors **a** and **b**,

$$\mathbf{a} \cdot (-\mathbf{b}) = -\mathbf{a} \cdot \mathbf{b}.$$

Proof. We have

$$\begin{aligned}\mathbf{a} \cdot (-\mathbf{b}) &= |\mathbf{a}|\ (\text{projection of } -\mathbf{b} \text{ on } \mathbf{a}),\\ &= |\mathbf{a}|\ (-1)\ (\text{projection of } \mathbf{b} \text{ on } \mathbf{a}),\\ &= -|\mathbf{a}|\ |\mathbf{b}| \cos\theta,\\ &= -\mathbf{a} \cdot \mathbf{b}.\end{aligned}$$

8·8·2. Properties of Scalar Product

Theorem 8·5. *The scalar product of two vectors is commutative, that is,*

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a},$$

for any two vectors **a** *and* **b**.

Proof. Let θ be the angle between **a** and **b**. Then,

$$\begin{aligned}\mathbf{a} \cdot \mathbf{b} &= |\mathbf{a}|\ |\mathbf{b}| \cos\theta,\\ &= |\mathbf{a}|\ |\mathbf{b}| \cos(-\theta),\\ &= \mathbf{b} \cdot \mathbf{a}.\end{aligned}$$

Theorem 8·6. (Distributive Law). *The scalar product distributes itself over addition, that is*

$$\mathbf{a} \cdot (\mathbf{b}+\mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c},$$

for any three vectors **a**, **b** *and* **c**.

Proof. Let the vector **a** lie along the line OX and let **a** be the unit vector along OX.

Let $\overrightarrow{LM}$ represent **b** and $\overrightarrow{MN}$ represents **c**, so that $\overrightarrow{LN}$ represents $\mathbf{b}+\mathbf{c}$. Let P, Q, R be the feet of perpendiculars on OX, drawn from the points L, M, and N respectively. Then

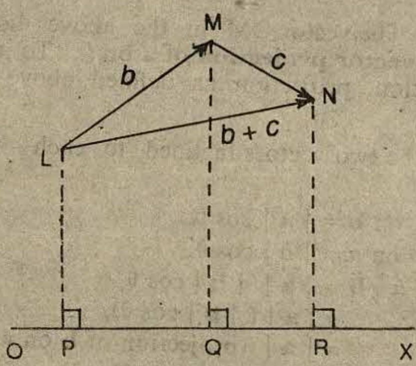


Fig. 8·21.

$\mathbf{a} \cdot (\mathbf{b}+\mathbf{c}) = |\mathbf{a}|$ (projection of $\mathbf{b}+\mathbf{c}$ on OX),
$= |\mathbf{a}|$ (projection of LN on OX),
$= |\mathbf{a}|$ (PR),
$= |\mathbf{a}|$ (PQ+QR),
$= |\mathbf{a}|$ (PQ)+ $|\mathbf{a}|$ (QR),
$= |\mathbf{a}|$ (projection of $\mathbf{b}$ on $\mathbf{a}$)
$+ |\mathbf{a}|$ (projection of $\mathbf{c}$ on $\mathbf{a}$),
$= \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$.

Corollary 1. If $\mathbf{a}$, $\mathbf{b}$ be any vectors, then

$$\mathbf{a} \cdot (-\mathbf{b}) = -(\mathbf{a} \cdot \mathbf{b}).$$

Proof. $\mathbf{a} \cdot (-\mathbf{b}) + \mathbf{a} \cdot \mathbf{b}$,
$= \mathbf{a} \cdot [(-\mathbf{b}) + \mathbf{b}]$,
$= \mathbf{a} \cdot \mathbf{0}$,
$= 0$.

Hence, $\mathbf{a} \cdot (-\mathbf{b}) = -(\mathbf{a} \cdot \mathbf{b})$.

Corollary 2. *If* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ *be any vectors, then*

$$\mathbf{a} \cdot (\mathbf{b}-\mathbf{c}) = \mathbf{a} \cdot \mathbf{b} - \mathbf{a} \cdot \mathbf{c}.$$

Theorem 8.7. *If* $\mathbf{a}$ *and* $\mathbf{b}$ *are any two vectors, and* k *a scalar, then*

$$(k\mathbf{a}) \cdot \mathbf{b} = k(\mathbf{a} \cdot \mathbf{b}) = \mathbf{a} \cdot (k\mathbf{b}).$$

Proof. Three different cases arise :

Case (*i*) $k=0$.

$$(k\mathbf{a}) \cdot \mathbf{b} = k(\mathbf{a} \cdot \mathbf{b}) = \mathbf{a} \cdot (k\mathbf{b}) = 0,$$

because $k\mathbf{a}$, $k\mathbf{b}$ are both zero vectors.

Case (*ii*) $k>0$.

Since $k>0$, therefore $\mathbf{a}$ and $k\mathbf{a}$ have the same direction and the magnitude of $k\mathbf{a}$ is $k \mid \mathbf{a} \mid$; also the direction of $k\mathbf{a}$ is the same as that of $\mathbf{a}$, so that if θ be the angle between $\mathbf{a}$ and $\mathbf{b}$, then the angle between $k\mathbf{a}$ and $\mathbf{b}$ is also θ.

$$\begin{aligned}(k\mathbf{a}) \cdot \mathbf{b} &= \mid k\mathbf{a} \mid \ \mid \mathbf{b} \mid \cos\theta,\\ &=(k \mid \mathbf{a} \mid) \mid \mathbf{b} \mid \cos\theta,\\ &=k(\mid \mathbf{a} \mid \ \mid \mathbf{b} \mid \cos\theta),\\ &=k(\mathbf{a} \cdot \mathbf{b}).\end{aligned}$$

$$\begin{aligned}\mathbf{a} \cdot (k\mathbf{b}) &=(k\mathbf{b}) \cdot \mathbf{a}, && \text{(By commutativity)}\\ &=k(\mathbf{b} \cdot \mathbf{a}),\\ &=k(\mathbf{a} \cdot \mathbf{b}). && \text{(By commutativity)}\end{aligned}$$

Thus $(k\mathbf{a}) \cdot \mathbf{b}=k(\mathbf{a} \cdot \mathbf{b})=\mathbf{a} \cdot (k\mathbf{b})$.

Case *(iii)* $k<0$.

Since the vectors $k\mathbf{a}$, $-k\mathbf{a}$ are negatives of each other, therefore

$$\begin{aligned}(k\mathbf{a}) \cdot \mathbf{b} &=-[(-k\mathbf{a}) \cdot \mathbf{b}],\\ &=-[(-k)(\mathbf{a} \cdot \mathbf{b})], && \text{[by case (ii), since } -k>0]\\ &=k(\mathbf{a} \cdot \mathbf{b}).\end{aligned}$$

Also, $\quad \mathbf{a} \cdot (k\mathbf{b})=k(\mathbf{a} \cdot \mathbf{b}) \quad$ [as in case *(ii)* above]

Therefore $\quad (k\mathbf{a}) \cdot \mathbf{b}=k(\mathbf{a} \cdot \mathbf{b})=\mathbf{a} \cdot (k\mathbf{b})$.

8·8·3. Some Useful Identities

(i) $(\mathbf{a}+\mathbf{b})^2=\mathbf{a}^2+\mathbf{b}^2+2\mathbf{a} \cdot \mathbf{b}$

(ii) $(\mathbf{a}-\mathbf{b})^2=\mathbf{a}^2+\mathbf{b}^2-2\mathbf{a} \cdot \mathbf{b}$

(iii) $(\mathbf{a}+\mathbf{b}) \cdot (\mathbf{a}-\mathbf{b})=\mathbf{a}^2-\mathbf{b}^2$

(iv) $\mathbf{a} \cdot \mathbf{b}=\frac{1}{4}[(\mathbf{a}+\mathbf{b})^2-(\mathbf{a}-\mathbf{b})^2]$.

We shall prove only *(iii)*. The simple proofs of the others are left as exercises for the reader.

Proof of *(iii)*.

$$\begin{aligned}(\mathbf{a}+\mathbf{b}) \cdot (\mathbf{a}-\mathbf{b}) &=(\mathbf{a}+\mathbf{b}) \cdot \mathbf{a}-(\mathbf{a}+\mathbf{b}) \cdot \mathbf{b},\\ &=\mathbf{a} \cdot \mathbf{a}+\mathbf{b} \cdot \mathbf{a}-\mathbf{a} \cdot \mathbf{b}-\mathbf{b} \cdot \mathbf{b},\\ &=\mathbf{a} \cdot \mathbf{a}-\mathbf{b} \cdot \mathbf{b}, && \text{since } \mathbf{b} \cdot \mathbf{a}=\mathbf{a} \cdot \mathbf{b}\\ &=\mathbf{a}^2-\mathbf{b}^2.\end{aligned}$$

Example 6. *Show that the components of a vector* $\mathbf{b}$ *along and perpendicular to a vector* $\mathbf{a}$ *are*

$$\frac{(\mathbf{a} \cdot \mathbf{b})\,\mathbf{a}}{\mathbf{a} \cdot \mathbf{a}} \quad \textit{and} \quad \frac{(\mathbf{a} \cdot \mathbf{a})\,\mathbf{b}-(\mathbf{a} \cdot \mathbf{b})\,\mathbf{a}}{\mathbf{a} \cdot \mathbf{a}}$$

respectively.

Solution. Let $\overrightarrow{OA}=\mathbf{a}$ and $\overrightarrow{OB}=\mathbf{b}$. Let C be the foot of the perpendicular from B on OA. Then

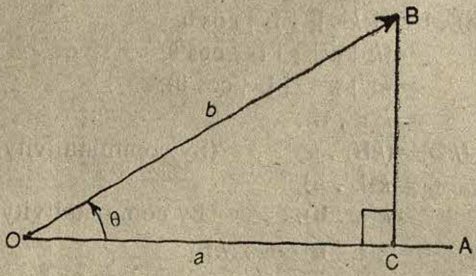


Fig. 8·22.

$$\mathbf{b}=\overrightarrow{OC}+\overrightarrow{CB}.$$

Therefore $\overrightarrow{OC}$ and $\overrightarrow{CB}$ are the components of $\mathbf{b}$ along and perpendicular to $\mathbf{a}$.

Now, $\overrightarrow{OC}=OC$ (unit vector along $\mathbf{a}$)

$$=OB\cos\theta\left(\frac{\mathbf{a}}{OA}\right),$$

$$=(OA\ .\ OB\cos\theta)\,\frac{\mathbf{a}}{(OA)^2},$$

$$=\frac{(\mathbf{a}\ .\ \mathbf{b})\,\mathbf{a}}{\mathbf{a}\ .\ \mathbf{a}},$$

since $(OA)^2=|\,\mathbf{a}\,|^2=\mathbf{a}\ .\ \mathbf{a}.$

Also, $\overrightarrow{CB}=\mathbf{b}-\dfrac{(\mathbf{a}\ .\ \mathbf{b})\,\mathbf{a}}{\mathbf{a}\ .\ \mathbf{a}},$

$$=[(\mathbf{a}\ .\ \mathbf{a})\,\mathbf{b}-(\mathbf{a}\ .\ \mathbf{b})\,\mathbf{a}]\ /\ (\mathbf{a}\ .\ \mathbf{a}).$$

8.8.4. Expression for Scalar Product in Terms of Components

As an application of the distributive law we can obtain a useful expression for the scalar product of two vectors in terms of their rectangular components.

Theorem 8·8. *Let*

$$\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k},$$
$$\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k}.$$

Then $\mathbf{a}\ .\ \mathbf{b}=a_1b_1+a_2b_2+a_3b_3.$

Proof. $\mathbf{a} \cdot \mathbf{b}=(a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}) \cdot (b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k})$,

$$=a_1\mathbf{i} \cdot (b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k})$$
$$+a_2\mathbf{j} \cdot (b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k})$$
$$+a_3\mathbf{k} \cdot (b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k}),$$
$$=a_1b_1\mathbf{i} \cdot \mathbf{i}+a_1b_2\mathbf{i} \cdot \mathbf{j}+a_1b_3\mathbf{i} \cdot \mathbf{k}$$
$$+a_2b_1\mathbf{j} \cdot \mathbf{i}+a_2b_2\mathbf{j} \cdot \mathbf{j}+a_2b_3\mathbf{j} \cdot \mathbf{k}$$
$$+a_3b_1\mathbf{k} \cdot \mathbf{i}+a_3b_2\mathbf{k} \cdot \mathbf{j}+a_3b_3\mathbf{k} \cdot \mathbf{k},$$
$$=a_1b_1+a_2b_2+a_3b_3,$$

since $\mathbf{i} \cdot \mathbf{i}=\mathbf{j} \cdot \mathbf{j}=\mathbf{k} \cdot \mathbf{k}=1$,

and $\mathbf{i} \cdot \mathbf{j}=\mathbf{j} \cdot \mathbf{i}=\mathbf{j} \cdot \mathbf{k}=\mathbf{k} \cdot \mathbf{j}=\mathbf{k} \cdot \mathbf{i}=\mathbf{i} \cdot \mathbf{k}=0$.

Corollary 1. *If the rectangular components of a vector* **a** *be* (a_1, a_2, a_3), *then*

$$|\mathbf{a}| = \sqrt{(a_1^2+a_2^2+a_3^2)}.$$

Proof. $\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$,

so that $\mathbf{a} \cdot \mathbf{a}=a_1^2+a_2^2+a_3^2$,

i.e., $|\mathbf{a}|^2=a_1^2+a_2^2+a_3^2$.

Hence $|\mathbf{a}| = \sqrt{(a_1^2+a_2^2+a_3^2)}$.

Corollary 2. *The angle* θ *between the vectors*

$$\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k},$$

and

$$\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k},$$

is given by

$$\cos\theta=\frac{a_1b_1+a_2b_2+a_3b_3}{\sqrt{(a_1^2+a_2^2+a_3^2)}\sqrt{(b_1^2+b_2^2+b_3^2)}}.$$

Proof. $\mathbf{a} \cdot \mathbf{b}=a_1b_1+a_2b_2+a_3b_3$, ...(1)

$|\mathbf{a}| =\sqrt{(a_1^2+a_2^2+a_3^2)}$, ...(2)

$|\mathbf{b}| =\sqrt{(b_1^2+b_2^2+b_3^2)}$. ...(3)

Also, $\mathbf{a} \cdot \mathbf{b}= |\mathbf{a}|\ |\mathbf{b}| \cos\theta$. ...(4)

Substituting the values of $\mathbf{a} \cdot \mathbf{b}$, $|\mathbf{a}|$, $|\mathbf{b}|$ in (4) we have the desired result.

Example 7. *Show that the points A, B, C whose position vectors are*

$\mathbf{a}=3\mathbf{i}-4\mathbf{j}-4\mathbf{k}$, $\mathbf{b}=2\mathbf{i}-\mathbf{j}+\mathbf{k}$,

$\mathbf{c}=\mathbf{i}-3\mathbf{j}-5\mathbf{k}$ *respectively are the verties of a right-angled triangle.*

Solution :

$$\overrightarrow{AB}=\mathbf{b}-\mathbf{a}=-\mathbf{i}+3\mathbf{j}+5\mathbf{k},$$

$$\overrightarrow{BC}=\mathbf{c}-\mathbf{b}=-\mathbf{i}-2\mathbf{j}-6\mathbf{k},$$

$\overrightarrow{CA}=\mathbf{a}-\mathbf{c}=2\mathbf{i}-\mathbf{j}+\mathbf{k}.$

$\overrightarrow{AB}.\overrightarrow{BC}=(-\mathbf{i}+3\mathbf{j}+5\mathbf{k}).(-\mathbf{i}-2\mathbf{j}-6\mathbf{k})$
$=(-1)(-1)+3(-2)+5(-6)=-35,$

$\overrightarrow{BC}.\overrightarrow{CA}=(-1).2+(-2)(-1)+(-6).1=-6$

$\overrightarrow{CA}.\overrightarrow{AB}=2(-1)+(-1).3+1.5=0.$

Since $\overrightarrow{CA}.\overrightarrow{AB}=0$, therefore $\angle BAC$ is a right angle.

Hence the triangle ABC is rt-angled at A,

EXERCISE 8 (*c*)

1. Find the scalar product of the vectors **a** and **b**, inclined to each other at an angle θ, if
   (*i*) $|\mathbf{a}|=2,\ |\mathbf{b}|=3,\ \theta=60°$,
   (*ii*) $|\mathbf{a}|=3,\ |\mathbf{b}|=4,\ \theta=135°$,
   (*iii*) $|\mathbf{a}|=4,\ |\mathbf{b}|=5,\ \theta=30°$.
2. Find the angle between **a** and **b** if $\mathbf{a}\ .\ \mathbf{b}=3,\ |\mathbf{a}|=\frac{3}{4},\ |\mathbf{b}|=8$.
3. The angle between two vectors **a** and **b** is 60°. If $\mathbf{a}.\mathbf{b}=4$, and $|\mathbf{a}|=|\mathbf{b}|$, find $|\mathbf{a}|$.
4. Find the scalar product of the following vectors :
   (*i*) $3\mathbf{i}+2\mathbf{j}-\mathbf{k}$ and $\mathbf{i}-\mathbf{j}+2\mathbf{k}$.
   (*ii*) $4\mathbf{i}-3\mathbf{j}+\mathbf{k}$ and $2\mathbf{i}+\mathbf{j}-\mathbf{k}$.
   (*iii*) $2\mathbf{i}+\mathbf{j}-3\mathbf{k}$ and $\mathbf{i}+\mathbf{j}-4\mathbf{k}$.
5. Find the magnitude of each of the following vectors :
   (*i*) $2\mathbf{i}+3\mathbf{j}+6\mathbf{k}$.
   (*ii*) $4\mathbf{i}+\mathbf{j}-8\mathbf{k}$.
6. Find the angle between the following pairs of vectors :
   (*i*) $\mathbf{a}=\mathbf{i}+2\mathbf{j}-2\mathbf{k},\ \mathbf{b}=9\mathbf{i}+\mathbf{j}+2\mathbf{k}$.
   (*ii*) $\mathbf{b}=2\mathbf{i}-\mathbf{j}+3\mathbf{k},\ \mathbf{b}=-4\mathbf{i}+2\mathbf{j}-6\mathbf{k}$.
7. Show that the vectors $2\mathbf{i}-3\mathbf{j}+5\mathbf{k}$ and $4\mathbf{i}+6\mathbf{j}+2\mathbf{k}$ are prependicular.
8. Show that if $|\mathbf{a}+\mathbf{b}|=|\mathbf{a}-\mathbf{b}|$, then the vectors **a** and **b** must be prependicular.
9. If **a** and **b** are prependicular vectors, show that
   $$|\mathbf{a}+\mathbf{b}|^2=|\mathbf{a}-\mathbf{b}|^2.$$
10. Show that the vectors **a** and **b** are prependicular if and only if
   $$|\mathbf{a}+\mathbf{b}|^2=|\mathbf{a}|^2+|\mathbf{b}|^2.$$

11. Show that for any pair of vectors **a** and **b**,
$$|\mathbf{a}+\mathbf{b}| \leqslant |\mathbf{a}| + |\mathbf{b}|.$$

12. If **a** and **b** be both unit vectors inclined to each other at an angle θ, then show that
$$|\mathbf{a}-\mathbf{b}| = 2 \sin\frac{\theta}{2}.$$

13. If **a**, **b**, **c** are vectors such that $\mathbf{a}+\mathbf{b}+\mathbf{c}=\mathbf{O}$, show that $\mathbf{c}^2=\mathbf{a}^2+\mathbf{b}^2+2\mathbf{a}.\mathbf{b}$. Interpret the result geometrically.

8·9. APPLICATIONS OF SCALAR PRODUCT TO GEOMETRY AND MECHANICS

We shall now consider some simple applications of scalar product to geometry and mechanics.

8·9·1. Applications to geometry

We have already seen how vector methods can be useful in proving geometrical results. We shall now study some more examples in which we shall use the concept of scalar product, specially to establish perpendicularity of certain lines.

Example 8. (*Cosine formula for triangles*). *prove that for any triangle ABC,*
$$a^2=b^2+c^2-2bc \cos A,$$
where the symbols have their usual meaning.

Solution :

Since $\quad \vec{BC}=\vec{BA}+\vec{AC},$

therefore $\quad \vec{BC}.\vec{BC}=(\vec{BA}+\vec{AC}).(\vec{BA}+\vec{AC})$

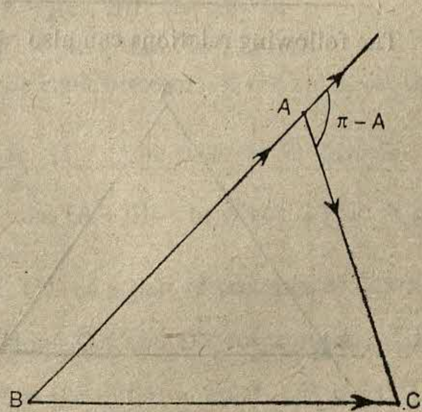


Fig. 8·23.

or $$BC^2=BA^2+AC^2+2\vec{BA}.\vec{AC}$$

or $$a^2=c^2+b^2+2cb\cos(\pi-A)$$

or $$\boxed{a^2=b^2+c^2-2\,bc\cos A}$$

Remark. The following relations can also be proved in a similar way :

$$b^2=c^2+a^2-2\,ca\cos B,$$
$$c^2=a^2+b^2-2\,ab\cos C.$$

Example 9. (*Projection formula for triangles*). *Prove that for any triangle ABC,*

$$c=b\cos A+a\cos B$$

Solution.

Since $\vec{AC}+\vec{CB}=\vec{AB}$,

therefore

$$\vec{AC}.\vec{AB}+\vec{CB}.\vec{AB}=\vec{AB}.\vec{AB}$$

or $$bc\cos A+ac\cos B=c^2$$

or $$b\cos A+a\cos B=c,$$ since $c\neq 0$

Therefore

$$\boxed{c=b\cos A+a\cos B}$$

Remark. The following relations can also be proved in a

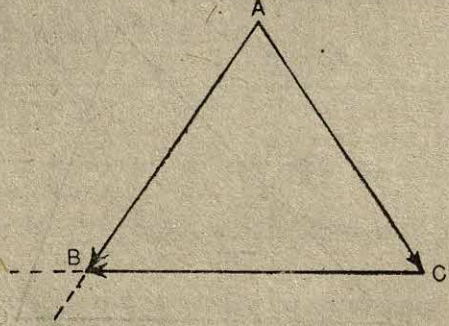


Fig. 8·24.

similar way :

$$\boxed{\begin{aligned} a&=b\cos C+c\cos B\\ b&=c\cos A+a\cos C\end{aligned}}$$

Example 10. *Prove that the right bisectors of the sides of a triangle are concurrent.*

Solution.

Let the position vectors of the vertices of a triangle ABC with respect to any origin of reference O be $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ respectively. Then the position vectors of the mid-points of BC, CA and AB are $\frac{1}{2}(\mathbf{b}+\mathbf{c})$, $\frac{1}{2}(\mathbf{c}+\mathbf{a})$, $\frac{1}{2}(\mathbf{a}+\mathbf{b})$ respectively. Let the perpendicular bisectors of AB and AC meet at H, and let the position vector of H be $\mathbf{h}$. Since $HF \perp AB$, we have

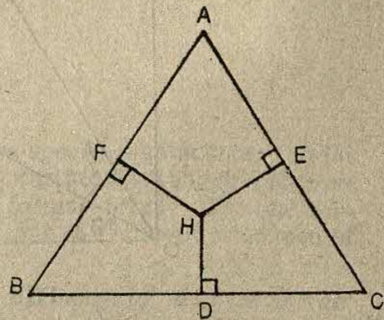


Fig. 8·25.

$$[\mathbf{h}-\tfrac{1}{2}(\mathbf{a}+\mathbf{b})].(\mathbf{b}+\mathbf{a})=0$$

$$\mathbf{h}.(\mathbf{b}-\mathbf{a})=\tfrac{1}{2}(\mathbf{a}+\mathbf{b}).(\mathbf{b}-\mathbf{a}) \qquad ...(1)$$

Also, since $HE \perp CA$, therefore we have

$$\mathbf{h}.(\mathbf{a}-\mathbf{c})=\tfrac{1}{2}(\mathbf{c}+\mathbf{a}).(\mathbf{a}-\mathbf{c}) \qquad ...(2)$$

Adding (1) and (2) throughout we have

$$\mathbf{h}.(\mathbf{b}-\mathbf{c})=\tfrac{1}{2}(\mathbf{c}+\mathbf{b}).(\mathbf{b}-\mathbf{c})$$

or $$[\mathbf{h}-\tfrac{1}{2}(\mathbf{c}+\mathbf{b})]\,.\,(\mathbf{b}-\mathbf{c})=0. \qquad ...(3)$$

From (3) we find that the $HD \perp BC$.

Hence the right bisectors of the sides of the triangle ABC are concurrent.

Example 11. (*Sum theorem for cosines*). *Prove that for any angles A and B,*

$$\cos(A-B)=\cos A\cos B+\sin A\sin B.$$

Solution.

Let OX, OY be a pair of rectangular axes in the plane, $\mathbf{i}, \mathbf{j}$ unit vectors along OX and OY respectively, $\vec{OP}, \vec{OQ}$ two unit vectors having O as the initial point. If $\vec{OP}$ and $\vec{OQ}$ make angles B and A respectively with OX, then $\angle POQ=A-B$.

The positions vectors **u** and **v** of P and Q are given by

$$\mathbf{u}=(\cos B)\mathbf{i}+(\sin B)\mathbf{j}.$$
$$\mathbf{v}=(\cos A)\mathbf{i}+(\sin A)\mathbf{j}.$$

Taking scalar product of **u** and **v**, we have

$$\mathbf{u}.\mathbf{v}=\cos A \text{ and } \cos B+\sin A \sin B, \quad ...(1)$$

since $\mathbf{i}.\mathbf{i}=\mathbf{j}.\mathbf{j}=1, \mathbf{i}.\mathbf{j}=0.$

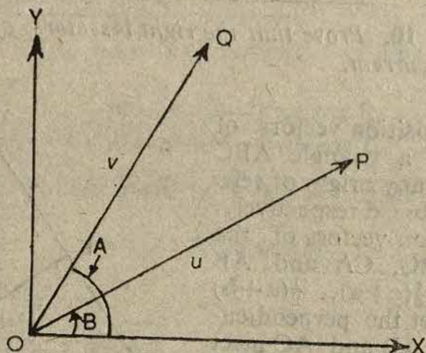


Fig. 8·26.

But $\mathbf{u}.\mathbf{v}=|\mathbf{u}|.|\mathbf{v}| \cos(A-B)=\cos(A-B), \quad ...(2)$

since $|\mathbf{u}|=|\mathbf{v}|=1.$

From (1) and (2) we find that

$$\boxed{\cos(A-B)=\cos A \cos B+\sin A \sin B}$$

Example 12. *Prove that angle in a semi-circle is a right angle.*

Solution. Let ABC be a semi-circle whose bounding diameter is AB, centre is O, and radius is r.

Take O as the origin of reference. Since the vectors $\vec{OA}, \vec{OB}$ have the same magnitude and support, but have equal senses,

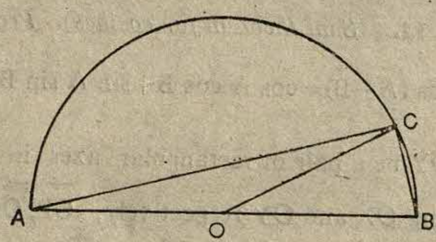


Fig. 8·27.

therefore,

$$\vec{OA}+\vec{OB}=\mathbf{0}.$$

Therefore if the position vector of A is $\mathbf{a}$, the position vector of B is $-\mathbf{a}$. Let the position vector of C be $\mathbf{c}$.

Since $\quad OC=OA=r$, therefore

$$OC^2=OA^2$$

or $\quad \mathbf{c}^2=\mathbf{a}^2$

or $\quad (\mathbf{c}-\mathbf{a}).(\mathbf{c}+\mathbf{a})=0$

or $\quad \vec{AC}\,.\,\vec{BC}=0,$

showing that $\angle ACB$ is a right angle.

Remark. The above result has a special significance. So far as the recorded history goes, this result can be called the *first theorem* of mathematics which was given a formal proof. The one who proved it, was Thales (624—548 B.C.), one of the seven wise men of antiquity.

Example 13. *Show that the altitudes of a triangle are concurrent.*

Solution. Let ABC be a triangle, and let the altitudes BM and CN meet in O. Join AO and produce it to meet BC in L. We shall show that AL is perpendicular to BC. Let us take O as the origin of reference, and let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ be the position vectors of A, B and C respectively. Then $\vec{AC}=\vec{OC}-\vec{OA}=\mathbf{c}-\mathbf{a}$, and $\vec{OM}=t\mathbf{b}$ for

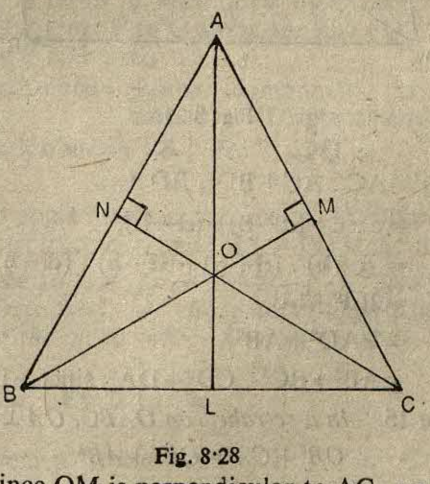


Fig. 8·28

some scalar t. Since OM is perpendicular to AC, we have

$$(t\mathbf{b})\,.\,(\mathbf{c}-\mathbf{a})=0 \text{ or } \mathbf{b}\,.\,\mathbf{c}=\mathbf{b}\,.\,\mathbf{a}. \qquad ...(1)$$

Similarly, using the fact that ON is perpendicular to AB, we have $\mathbf{c} \cdot \mathbf{b}=\mathbf{c} \cdot \mathbf{a}$. ...(2)

From (1) and (2), we have $\mathbf{b} \cdot \mathbf{a}=\mathbf{c} \cdot \mathbf{a}$, that is $(\mathbf{b}-\mathbf{c}) \cdot \mathbf{a}=0$, showing that AL is perpendicular to BC, *i.e.*, AL is an altitude of the triangle ABC.

Hence the altitudes AL, BM and CN all pass through O, and are therefore concurrent.

Example 14. *Show that the sum of the squares on the sides of a parallelogram is equal to the sum of the squares on the diagonals.*

Solution. Let ABCD be a parallelogram. Take A as the origin of reference. Let **b**, **c** and **d** be the position vectors of the vertices B, C and D respectively.

$$\vec{AC}=\vec{AD}+\vec{DC}=\vec{AD}+\vec{AB}.$$

$$\therefore \qquad \mathbf{c}=\mathbf{d}+\mathbf{b}$$

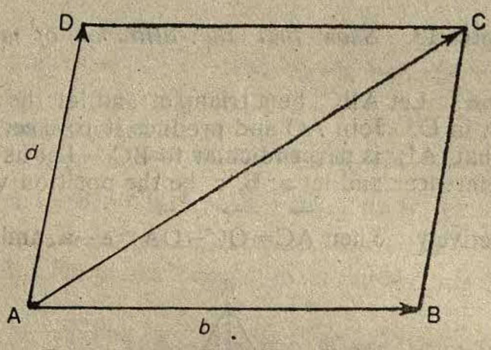


Fig. 8·29.

$$\begin{aligned}AC^2+BD^2&=\vec{AC} \cdot \vec{AC}+\vec{BD} \cdot \vec{BD}\\&=\mathbf{c} \cdot \mathbf{c}+(\mathbf{d}-\mathbf{b}) \cdot (\mathbf{d}-\mathbf{b}),\\&=\mathbf{d}+\mathbf{b}) \cdot (\mathbf{d}+\mathbf{b})+(\mathbf{d}-\mathbf{b}) \cdot (\mathbf{d}-\mathbf{b}),\\&=2(\mathbf{d}^2+\mathbf{b}^2),\\&=2(AD^2+AB^2),\\&=AB^2+BC^2+CD^2+DA^2, \text{ since } AD=BC,\ AB=CD.\end{aligned}$$

Example 15. *In a tetrahedron OABC, OA ⊥ BC. Show that*

$$OB^2+CA^2=OC^2+AB^2.$$

Solution. Take O as the origin of reference. Let the position vectors of A, B, C be **a**, **b**, **c** respectively.

Since $OA \perp BC$, therefore $\overrightarrow{OA} \cdot \overrightarrow{BC}=0$, *i.e.*, $\mathbf{a} \cdot (\mathbf{c}-\mathbf{b})=0$, or $\mathbf{a} \cdot \mathbf{c}=\mathbf{a} \cdot \mathbf{b}$. ...(1)

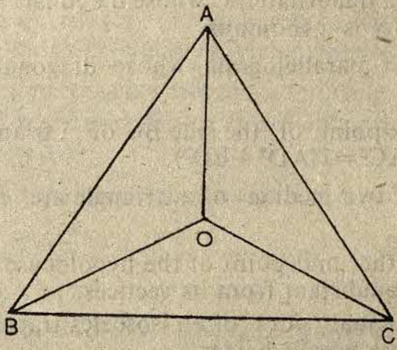


Fig. 8·30.

Now $OB^2+CA^2=\mathbf{b}^2+(\mathbf{a}-\mathbf{c})^2$,

$$=\mathbf{b}^2+\mathbf{a}^2-2\mathbf{a} \cdot \mathbf{c}+\mathbf{c}^2,$$

$$=\mathbf{b}^2+\mathbf{a}^2-2\mathbf{a} \cdot \mathbf{b}+\mathbf{c}^2, \text{ by using (1)}$$

$$=\mathbf{c}^2+(\mathbf{a}^2-2\mathbf{a} \cdot \mathbf{b}+\mathbf{b}^2),$$

$$=\mathbf{c}^2+(\mathbf{a}-\mathbf{b})^2,$$

$$=OC^2+AB^2.$$

8·9·2. Work Done by a Force

The scalar product of two vectors can be given a physical interpretation in terms of work done by a force when it acts on an object moving it from P to Q along the line PQ. Work done by a constant force F is defined as

Work done=(Force component in the direction of motion). (Displacement)

$\therefore$ Work done$=(\,|\,\mathbf{F}\,|\,\cos\theta) \,.\, PQ$

$$=\mathbf{F} \cdot \mathbf{d},$$

where θ is the angle between the line of action of the force and the direction of displacement, and $|\,\mathbf{d}\,|=PQ$ is the displacement vector.

Example 16. *A force* $\mathbf{F}=2\mathbf{i}+\mathbf{j}-\mathbf{k}$ *acts at a point A whose position vector is* $2\mathbf{i}-\mathbf{j}$. *If the point of application of* $\mathbf{F}$ *moves from the point A to the point whose position vector is* $2\mathbf{i}+\mathbf{j}$, *find the work done by* $\mathbf{F}$.

Solution. $\mathbf{F}=2\mathbf{i}+\mathbf{j}-\mathbf{k}$.

Displacement vector$=\overrightarrow{AB}=(2\mathbf{i}+\mathbf{j})-(2\mathbf{i}-\mathbf{j})=2\mathbf{j}$.

$\therefore$ Work done $=\mathbf{F} \cdot \mathbf{d}=(2\mathbf{i}+\mathbf{j}-\mathbf{k}) \cdot 2\mathbf{j}$,

$$=2.$$

EXERCISE 8 (*d*)

1. Show that the diagonals of a rhombus are at right angles.
2. Show that a quadrilateral whose diagonals bisect each other at right angles is a rhombus.
3. Show that a parallelogram whose diagonals are equal is a rectangle.
4. D is the mid-point of the side BC of a triangle ABC. Show that $AB^2+AC^2=2(AD^2+BD^2)$.
5. Show that if two medians of a triangle are equal, the triangle is isoceles.
6. Show that the mid-point of the hypotenuse of a right-angled triangle is equidistant from its vectices.
7. Show that equal sides of an isosceles trapezium are equally inclined to the paralled sides.
8. Show that the diagonals of an isosceles trapezium are equal.
9. If two pairs of opposite sides of a tetrahedron are at right angles, then show that the third pair is also at right angles.
10. If two opposite edges of a tetrahedron are equal in length and are at right angles to the line joining their mid-points, show that the remaining pairs of opposite edges have the same property.
11. The line joining the mid-points of two opposite edges of a tetrahedron is perpendicular to the edges. Show that the remaining pairs of opposite edges are equal.
12. Two opposite edges BC, AD of a tetrahedron are perpendicular to each other. Show that the distance between the mid-points of AB and CD is equal to the distance between the mid-points of AC and BD.
13. Constant forces $\mathbf{P}_1=\mathbf{i}-\mathbf{j}+\mathbf{k}$, $\mathbf{P}_2=-\mathbf{i}+2\mathbf{j}-\mathbf{k}$ and $\mathbf{P}_3=\mathbf{j}-\mathbf{k}$ act on a particle at a point A. Determine the work done when the particle is displaced from the position A to B, where $\mathbf{A}=4\mathbf{i}-3\mathbf{j}-2\mathbf{k}$ and $\mathbf{B}=6\mathbf{i}+\mathbf{j}-3\mathbf{k}$ are the position vectors of A and B. (*Roorkee Entrance, 1980*)
14. Constant forces $P=2\mathbf{i}-5\mathbf{j}+9\mathbf{k}$ and $Q=-\mathbf{i}+2\mathbf{j}-\mathbf{k}$ act on a particle. Determine the work done when the particle is displaced from a point A with position vector $4\mathbf{i}-3\mathbf{j}-2\mathbf{k}$ to a point B with position vector $6\mathbf{i}+\mathbf{j}-3\mathbf{k}$.
(*Roorkee Entrance: 1984*)

8·10. VECTOR PRODUCT OF TWO VECTORS

Definition 8·8. *The vector product of two vectors* **a** *and* **b** *to be denoted by* $\mathbf{a}\times\mathbf{b}$ *is the vector*

$$\mathbf{a}\times\mathbf{b}=(|\mathbf{a}|\,.\,|\mathbf{b}|\ \sin\theta)\,\mathbf{n},$$

where θ *is the angle between* **a** *and* **b**, **n** *a unit vector perpendicular to both* **a** *and* **b**, *the sense of* **n** *being such that* **a**, **b**, **n** *form a right-handed system.*

Since we use a cross ($\times$) to indicate the vector product of two vectors, therefore vector product is often called *cross product.*

The following particular cases are interesting :

(*i*) If **a** and **b** are parallel, so that $\theta=0$ or π, then $\mathbf{a}\times\mathbf{b}=0$.

(*ii*) If **a** and **b** are perpendicular, so that

$$\theta=\frac{\pi}{2},$$

then $\mathbf{a}\times\mathbf{b}=(\,|\mathbf{a}|\,.\,|\mathbf{b}|\,)\,\mathbf{n}$,

where **a**, **b**, **n** form a right-handed system,

(*iii*) For any vector **a**,

$$\mathbf{a}\times\mathbf{a}=0.$$

(*iv*) If **i**, **j**, **k** be a set of mutually perpendicular unit vectors forming a right-handed system, then

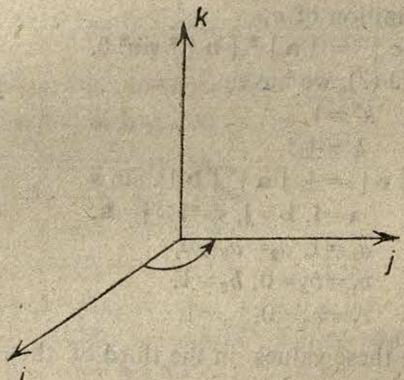


Fig. 8·31.

$$\mathbf{i}\times\mathbf{i}=0,\ \mathbf{j}\times\mathbf{j}=0,\ \mathbf{k}\times\mathbf{k}=0,$$
$$\mathbf{i}\times\mathbf{j}=\mathbf{k},\ \mathbf{j}\times\mathbf{k}=\mathbf{i},\ \mathbf{k}\times\mathbf{i}=\mathbf{j},$$
$$\mathbf{j}\times\mathbf{i}=-\mathbf{k},\ \mathbf{k}\times\mathbf{j}=-\mathbf{i},\ \mathbf{i}\times\mathbf{k}=-\mathbf{j}.$$

8·10·1. Expression for Vector Product in Terms of Components

Theorem 8·9. If $\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$,

$$\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k},$$

then

$$\mathbf{a}\times\mathbf{b}=(a_2b_3-a_3b_2)\,\mathbf{i}+(a_3b_1-a_1b_3)\,\mathbf{j}+(a_1b_2-a_2b_1)\,\mathbf{k}.$$

Proof. Let $\mathbf{c}=\mathbf{a}\times\mathbf{b}$,

and $\mathbf{c}=c_1\mathbf{i}+c_2\mathbf{j}+c_3\mathbf{k}$.

Since $\mathbf{c}$ is perpendicular to $\mathbf{a}$ as well as to $\mathbf{b}$, therefore,

$$c_1a_1+c_2a_2+c_3a_3=0,$$

and $$c_1b_1+c_2b_2+c_3b_3=0.$$

Solving the above equations for c_1, c_2, c_3, we have

$$\frac{c_1}{a_2b_3-a_3b_2}=\frac{c_2}{a_3b_1-a_1b_3}=\frac{c_3}{a_1b_2-a_2b_1}=k \text{ (say)},$$

i.e., $$\left.\begin{aligned} c_1&=k(a_2b_3-a_3b_2),\\ c_2&=k(a_3b_1-a_1b_3),\\ c_3&=k(a_1b_2-a_2b_1).\end{aligned}\right\} \qquad ...(1)$$

Squaring and adding, we have

$$\begin{aligned} c_1^2+c_2^2+c_3^2&=k^2\{(a_2b_3-a_3b_2)^2+(a_3b_1-a_1b_3)^2+(a_1b_2-a_2b_1)^2\}\\ &=k^2\{(a_1^2+a_2^2+a_3^2)\,(b_1^2+b_2^2+b_3^2)\\ &\qquad -(a_1b_1+a_2b_2+a_3b_3)^2\},\end{aligned}$$

or $$\begin{aligned} |\mathbf{c}|^2&=k^2\{|\mathbf{a}|^2|\mathbf{b}|^2-|\mathbf{a}|^2|\mathbf{b}|^2\cos^2\theta\}\\ &=k^2|\mathbf{a}|^2|\mathbf{b}|^2\sin^2\theta, \qquad ...(2)\end{aligned}$$

where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$.

Also, by definition of $\mathbf{c}$,

$$|\mathbf{c}|^2=|\mathbf{a}|^2|\mathbf{b}|^2\sin^2\theta. \qquad ...(3)$$

From (2) and (3), we have

$$k^2=1,$$

i.e., $$k=\pm 1,$$

so that $$|\mathbf{c}|=\pm|\mathbf{a}|\,|\mathbf{b}|.\sin\theta \qquad ...(4)$$

Letting $\mathbf{a}=\mathbf{i}$, $\mathbf{b}=\mathbf{j}$, $\mathbf{c}=\mathbf{i}\times\mathbf{j}=\mathbf{k}$,

$$a_1=1,\ a_2=a_3=0,$$
$$b_1=b_3=0,\ b_2=1,$$
$$c_1=c_2=0,\ c_3=1.$$

Substituting these values in the third of the equations (1), we have

$$1=k(1.1-0.0),$$

i.e., $$k=1. \qquad ...(5)$$

From (1) and (5), we have

$$c_1=a_2b_3-a_3b_2,$$
$$c_2=a_3b_1-a_1b_3,$$
$$c_3=a_1b_2-a_2b_1.$$

Hence $$\boxed{\begin{aligned}\mathbf{a}\times\mathbf{b}=(a_2b_3-a_3b_2)\mathbf{i}+(a_3b_1-a_1b_3)\mathbf{j}\\ +(a_1b_2-a_2b_1)\mathbf{k}\end{aligned}}$$

8·10 2. Properties of the Vector Product

Theorem 8·10. *Vector product is anti-commutative, that is, for any two vectors* $\mathbf{a}$ *and* $\mathbf{b}$,

$$\mathbf{a}\times\mathbf{b}=-(\mathbf{b}\times\mathbf{a}).$$

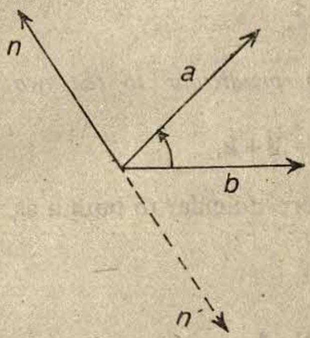


Fig. 8·32.

Proof. By definition,

$$\mathbf{a}\times\mathbf{b}=|\,\mathbf{a}\,|\;|\,\mathbf{b}\,|\;\sin\theta\;\mathbf{n},$$

and $$\mathbf{b}\times\mathbf{a}=|\,\mathbf{b}\,|\;|\,\mathbf{a}\,|\;\sin\theta\;\mathbf{n}',$$

The vectors $\mathbf{n}$ and $\mathbf{n}'$ are both perpendicular to the plane of $\mathbf{a}$ and $\mathbf{b}$. If $\mathbf{n}$ points upwards, then $\mathbf{n}'$ points downwards.

Hence $\mathbf{a}\times\mathbf{b}=-(\mathbf{b}\times\mathbf{a})$.

Remark. Note that in a vector product the order of vectors is important.

Theorem 8·11. *Vector product distributes itself over addition, that is,*

$$\mathbf{a}\times(\mathbf{b}+\mathbf{c})=\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c},$$

for every triple of vectors $\mathbf{a}$, $\mathbf{b}$ *and* $\mathbf{c}$.

Proof. Let
$$\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k},$$
$$\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k},$$
$$\mathbf{c}=c_1\mathbf{i}+c_2\mathbf{j}+c_3\mathbf{k},$$

Then $$\mathbf{b}+\mathbf{c}=(b_1+c_1)\,\mathbf{i}+(b_2+c_2)\,\mathbf{j}+(b_3+c_3)\,\mathbf{k}.$$

$$\therefore\quad \mathbf{a}\times(\mathbf{b}+\mathbf{c})=\Sigma[a_2(b_3+c_3)-a_3(b_2+c_2)]\,\mathbf{i},$$
$$=\Sigma(a_2b_3-a_3b_2)\,\mathbf{i}+\Sigma(a_2c_3-a_3c_2)\,\mathbf{i},$$
$$=\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}.$$

Theorem 8·12. *If* $\mathbf{a}$ *and* $\mathbf{b}$ *are any vectors, and s a scalar, then*

$$(s\mathbf{a})\times\mathbf{b}=s(\mathbf{a}\times\mathbf{b})=\mathbf{a}\times(s\mathbf{b}).$$

Proof. Let $\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$, $\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k}$,

Then $$s\mathbf{a}=sa_1\mathbf{i}+sa_2\mathbf{j}+sa_3k.$$

$$\therefore\quad (s\mathbf{a})\times\mathbf{b}=[(sa_2)b_3-(sa_3)b_2]\mathbf{i}$$
$$+[(sa_3)b_1-(sa_1)b_3]\mathbf{j}$$
$$+[(sa_1)b_2-(sa_2)b_1]\mathbf{k}.$$
$$=s(a_2b_3-a_2b_3)\mathbf{i}+s(a_3b_1-a_1b_3)\mathbf{j}$$
$$+s(a_1b_2-a_2b_1)\mathbf{k},$$

$$=s\Sigma(a_2b_3-a_3b_2)\mathbf{i},$$
$$=s(\mathbf{a}\times\mathbf{b}).$$

Also $\quad \mathbf{a}\times(s\mathbf{b})=-[(s\mathbf{b})\times\mathbf{a}],$
$$=-[s(\mathbf{b}\times\mathbf{a})],$$
$$=-[-s(\mathbf{a}\times\mathbf{b})],$$
$$=s(\mathbf{a}\times\mathbf{b}).$$

Example 17. *Find a unit vector perpendicular to the two vectors* $\mathbf{i}+2\mathbf{j}-\mathbf{k}$ *and* $2\mathbf{i}+3\mathbf{j}+\mathbf{k}$.

Solution. Let $\mathbf{a}=\mathbf{i}+2\mathbf{j}-\mathbf{k}$, $\mathbf{b}=2\mathbf{i}+3\mathbf{j}+\mathbf{k}$.

The vector $\dfrac{\mathbf{a}\times\mathbf{b}}{|\mathbf{a}\times\mathbf{b}|}$ is a unit vector perpendicular to both **a** as well as **b**.

Now $\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ 1 & 2 & -1\\ 2 & 3 & 1\end{vmatrix}=5\mathbf{i}-3\mathbf{j}-\mathbf{k},$

Also $\quad |\mathbf{a}\times\mathbf{b}|=\sqrt{\{5^2+(-3)^2+(-1)^2\}}=\sqrt{35}.$

A unit vector perpendicular to both **a** as well as **b** is given by
$$\frac{1}{\sqrt{35}}(5\mathbf{i}-3\mathbf{j}-\mathbf{k}).$$

Example 18. *Solve for* **r** :
$$\mathbf{r}\times\mathbf{a}=\mathbf{a}\times\mathbf{b}.$$

Solution. The given equation can be written as
$$\mathbf{r}\times\mathbf{a}-\mathbf{a}\times\mathbf{b}=\mathbf{0}$$
$$\Rightarrow \quad \mathbf{r}\times\mathbf{a}+\mathbf{b}\times\mathbf{a}=0$$
$$\Rightarrow \quad (\mathbf{r}+\mathbf{b})\times\mathbf{a}=0$$
$\therefore \quad \mathbf{r}+\mathbf{b}$ is parallel to **a**,

so that $\mathbf{r}+\mathbf{b}=t\,\mathbf{a}$, where $t\in\mathbf{R}$.

Thus $\quad \mathbf{r}=t\mathbf{a}-\mathbf{b}$ is the required solution, where t is a parameter.

Example 19. *Show that the three points whose position vectors are* $2\mathbf{a}+3\mathbf{b}+4\mathbf{c}$, $-7\mathbf{b}+10\mathbf{c}$ *and* $\mathbf{a}-2\mathbf{b}+3\mathbf{c}$ *are collinear.*

Solution. Let us label the points as A, B and C respectively. If 0 be the origin of reference, then
$$\overrightarrow{OA}=2\mathbf{a}+3\mathbf{b}-4\mathbf{c},$$
$$\overrightarrow{OB}=-7\mathbf{b}+10\mathbf{c},$$

$$\vec{OC}=\mathbf{a}-2\mathbf{b}+3\mathbf{c},$$

$$\vec{AB}=\vec{OB}-\vec{OA},$$
$$=(-7\mathbf{b}+10\mathbf{c})-(2\mathbf{a}+3\mathbf{b}-4\mathbf{c}),$$
$$=-2\mathbf{a}-10\mathbf{b}+14\mathbf{c}.$$

$$\vec{AC}=\vec{OC}-\vec{OA},$$
$$=-\mathbf{a}-5\mathbf{b}+7\mathbf{c}.$$

$$\vec{AB}\times\vec{AC}=(-2-\mathbf{a}-10\mathbf{b}+14\mathbf{c})\times(-\mathbf{a}-5\mathbf{b}+7\mathbf{c}),$$
$$=2(-\mathbf{a}-5\mathbf{b}+7\mathbf{c})\times(-\mathbf{a}-5\mathbf{b}+7\mathbf{c}),$$
$$=2\mathbf{p}\times\mathbf{p},\text{ where }\mathbf{p}=-\mathbf{a}-5\mathbf{b}+7\mathbf{c}$$
$$=\mathbf{0}.$$

Since $\vec{AB}\times\vec{AC}=\mathbf{0}$, therefore the vectors $\vec{AB}$ and $\vec{AC}$ are parallel. Since the parallel vectors $\vec{AB}$ and $\vec{AC}$ are co-initial, therefore the points A, B, C are collinear.

Remark. We could have observed that $\vec{AC}=2\vec{AB}$ and drawn the desired conclusion without taking vector products.

EXERCISE 8 (*e*)

1. If $\mathbf{a}=3\mathbf{i}+2\mathbf{j}-\mathbf{k}$, $\mathbf{b}=4\mathbf{i}+3\mathbf{j}+2\mathbf{k}$, find
   (*i*) $\mathbf{a}\times\mathbf{b}$, and (*ii*) $\mathbf{b}\times\mathbf{a}$.
2. If $\mathbf{a}=\mathbf{i}+3\mathbf{j}-2\mathbf{k}$, $\mathbf{b}=2\mathbf{i}-\mathbf{j}-\mathbf{k}$, $\mathbf{c}=2\mathbf{i}+3\mathbf{j}+4\mathbf{k}$, find
   (*i*) $\mathbf{a}\times\mathbf{b}$, (*iii*) $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$,
   (*ii*) $\mathbf{b}\times\mathbf{c}$, (*iv*) $(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$.
3. Find a unit vector perpendicular to both $\mathbf{a}=\mathbf{i}-2\mathbf{j}+\mathbf{k}$, and $\mathbf{b}=3\mathbf{i}+2\mathbf{j}-\mathbf{k}$.
4. If $\mathbf{a}$ and $\mathbf{b}$ are perpendicular vectors, and $|\mathbf{a}|=3$, $|\mathbf{b}|=2$, find $|\mathbf{a}\times\mathbf{b}|$.
5. If $|\mathbf{a}|=2$, $|\mathbf{b}|=4$, and the vectors $\mathbf{a}$ and $\mathbf{b}$ are inclined at an angle $30°$, find $|\mathbf{a}\times\mathbf{b}|$.
6. Show that for any vectors $\mathbf{a}$ and $\mathbf{b}$,
   $$(\mathbf{a}\times\mathbf{b})^2+(\mathbf{a}\cdot\mathbf{b})^2=|\mathbf{a}|^2|\mathbf{b}|^2.$$
7. Find a unit vector perpendicular to the plane of $\mathbf{a}$ and $\mathbf{b}$, where
   $\mathbf{a}=3\mathbf{i}+2\mathbf{j}+5\mathbf{k}$ and $\mathbf{b}=\mathbf{i}-3\mathbf{j}+\mathbf{k}$. (*A.I.S S.C.E., 1984*)

8. Find a unit vector perpendicular to the plane of **a** and **b** where
a=**i**−**j** and **b**=**j**+**k**. (*D.B.S.S.C.E., 1985*)

9. If $\mathbf{a}\times\mathbf{b}=\mathbf{b}\times\mathbf{c}\neq 0$, prove that $\mathbf{a}+\mathbf{c}=m\mathbf{b}$, where m is a scalar. (*A.I.S.S.C.E, 1985*)

10. **a**, **b**, **c** are three vectors such that $\mathbf{a}\times\mathbf{b}=\mathbf{c}$, $\mathbf{b}\times\mathbf{c}=\mathbf{a}$. Prove that **a**, **b**, **c** are mutually at right angles, and $|\mathbf{b}|=1$, $|\mathbf{c}|=|\mathbf{a}|$. (*A.I.S.S.C.E, 1986*)

11. Given $\mathbf{a}=\mathbf{i}-\mathbf{j}+2\mathbf{k}$, $\mathbf{b}=3\mathbf{i}+2\mathbf{j}-4\mathbf{k}$, and $\mathbf{c}=\mathbf{i}+\mathbf{j}+\mathbf{k}$, compute $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ and $(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$. Are these equal? (*A.I.S.S.C.E., 1986*)

12. If $\mathbf{a}+\mathbf{b}+\mathbf{c}=\mathbf{0}$, prove that
$\mathbf{a}\times\mathbf{b}=\mathbf{b}\times\mathbf{c}=\mathbf{c}\times\mathbf{a}$. (*A.I.S.S.C.E., 1986*)

13. Find unit vectors perpendiculars to the two vectors **i**+**j** and **j**+**k**. (*A.I.S.S.C.E., 1987*)

8·10·3. Applications of Vector Product to Geometry and Mechanics

Example 20. *Prove that the medians of triangle are concurrent.*

Solution. Let D, E, F be the mid-points of the sides BC, CA and AB respectively of a triangle ABC, and let the medians BE and CF intersect at G. We shall show that A, G, D are collinear.

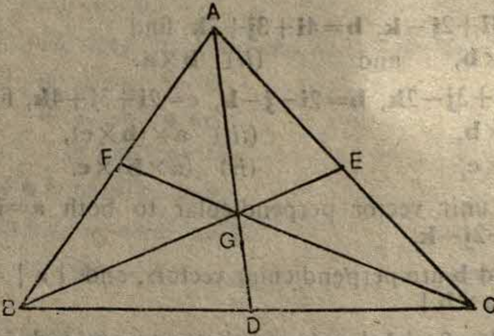


Fig. 8·33.

Join GA and GD. Let the position vectors of A, B and C with reference to G as the origin of reference be **a**, **b** and **c** respectively. Then

$$\vec{GE}=\vec{GC}+\vec{CE}=\vec{GC}+\tfrac{1}{2}\vec{CA},$$

$$=\vec{GC}+\tfrac{1}{2}(\vec{GA}-\vec{GC}),$$

$$=\tfrac{1}{2}(\overrightarrow{GC}+\overrightarrow{GA}),$$
$$=\tfrac{1}{2}(\mathbf{c}+\mathbf{a}).$$

Similarly,

$$\overrightarrow{GF}=\tfrac{1}{2}(\mathbf{a}+\mathbf{b}),\ \overrightarrow{GD}=\tfrac{1}{2}(\mathbf{b}+\mathbf{c}).$$

Since $\overrightarrow{GB}$ and $\overrightarrow{GE}$ are collinear,

we have $\quad \overrightarrow{GB}\times\overrightarrow{GE}=0,$

i.e., $\quad \mathbf{b}\times\tfrac{1}{2}(\mathbf{c}+\mathbf{a})=0,$

or $\quad \mathbf{b}\times\mathbf{c}+\mathbf{b}\times\mathbf{a}=0,$

or $\quad \mathbf{b}\times\mathbf{c}=\mathbf{a}\times\mathbf{b}.$ ...(1)

Since $\overrightarrow{GC}$ and $\overrightarrow{GF}$ are collinear,

we have

$$\mathbf{c}\times\mathbf{a}=\mathbf{b}\times\mathbf{c}. \qquad ...(2)$$

From (1) and (2), we have

$$\mathbf{c}\times\mathbf{a}=\mathbf{a}\times\mathbf{b},$$

or $\quad \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}=0,$

or $\quad \mathbf{a}\times(\mathbf{b}+\mathbf{c})=0,$

or $\quad \tfrac{1}{2}(\mathbf{b}+\mathbf{c})\times\mathbf{a}=0,$

or $\quad \overrightarrow{GD}\times\overrightarrow{GA}=0,$

showing that the vectors $\overrightarrow{GD}$ and $\overrightarrow{GA}$ are collinear, *i.e.*, the points G, D and A are collinear. This shows that the median AD passes through G. Thus the medians AD, BE and CF all pass through the point G.

Example 21. (*The law of sines*). *Prove that in any* $\triangle ABC$,

$$\frac{a}{\sin A}=\frac{b}{\sin B}=\frac{c}{\sin C}.$$

Solution. Let $\mathbf{a}=\overrightarrow{BC},\ \mathbf{b}=\overrightarrow{CA},\ \mathbf{c}=\overrightarrow{AB}.$

Then we know that

$$\mathbf{a}+\mathbf{b}+\mathbf{c}=0,$$

so that $\mathbf{a}=-(\mathbf{b}+\mathbf{c}).$

Taking cross product of both sides with $\mathbf{a}$, we have

$$\mathbf{a}\times\mathbf{a}=-(\mathbf{b}+\mathbf{c})\times\mathbf{a},$$

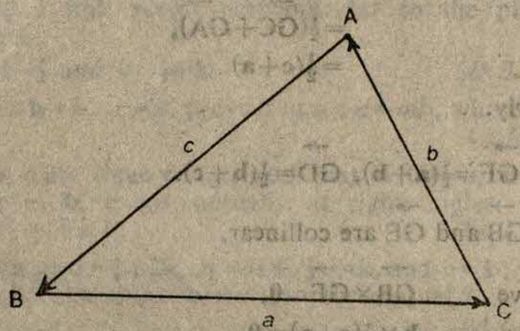


Fig. 8·34.

$$\text{or} \qquad 0=-(\mathbf{b}\times\mathbf{a})-(\mathbf{c}\times\mathbf{a}),$$

$$\text{or} \qquad \mathbf{c}\times\mathbf{a}=\mathbf{a}\times\mathbf{b}.$$

$$\therefore \qquad |\,\mathbf{c}\times\mathbf{a}\,| = |\,\mathbf{a}\times\mathbf{b}\,|,$$

i.e., $ac \sin B = ab \sin C$.

$$\therefore \qquad \frac{c}{\sin C}=\frac{b}{\sin B}\,. \qquad \text{...(1)}$$

$$\text{Similarly} \qquad \frac{a}{\sin A}=\frac{c}{\sin C}\,. \qquad \text{...(2)}$$

From (1) and (2), we have

$$\boxed{\frac{a}{\sin A}=\frac{b}{\sin B}=\frac{c}{\sin C}}$$

8·10·4. Interpretation of Vector Product as Vector Area

A plane area bounded by a closed curve which does not cross itself can be represented by a vector. In order to do so we have to distinguish between the two senses in which a curve can be described.

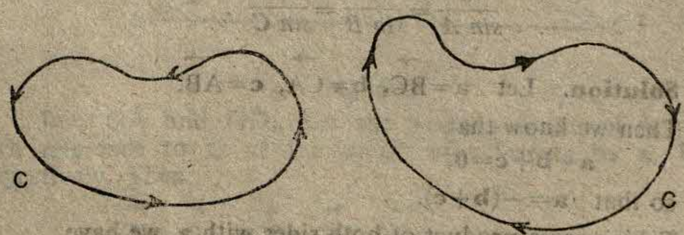


Fig. 8·35.

We can represent a plane area bounded by a closed curve C which does not cross itself by means of a vector **v** as follows :

(*i*) We take the magnitude of **v** to be the number of units of area bounded by C.

(*ii*) The support of **v** is taken to be perpendicular to the plane of C.

(*iii*) The sense of **v** is taken to be such that the sense of description of C and the sense of **v** correspond to a right-handed screw.

8·10·5. Area of a Triangle as a Vector Product

Let **a** and **b** two vectors having a common initial point O, and terminal points A and B respectively, so that $\vec{OA}=\mathbf{a}$, $\vec{OB}=\mathbf{b}$.

We shall find the vector **v** representing area of the triangle OAB. It will be shown that

$$\mathbf{v}=\tfrac{1}{2}\mathbf{a}\times\mathbf{b}. \qquad ...(1)$$

In fact, the magnitude of each side of (1) is $\frac{1}{2}$ OA . OB. sin θ, where θ is the angle AOB, and the support of each of them is the same.

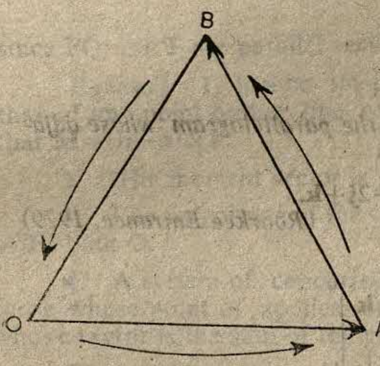


Fig. 8·36 (*a*)

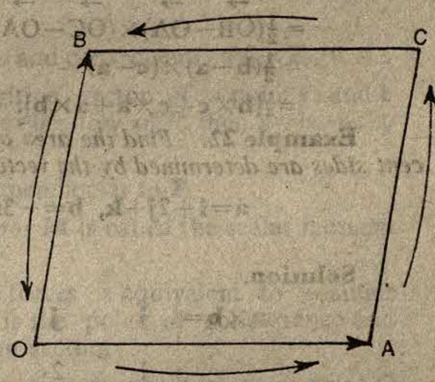


Fig. 8·36 (*b*)

Also, according to the convention explained above, senses of the vectors **v** and $\frac{1}{2}$ **a**$\times$**b** are the same.

Hence $\quad \mathbf{v}=\frac{1}{2}\,\mathbf{a}\times\mathbf{b}.$

Corollaries 1. Area of a parallelogram as a vector product. The vector area of the parallelogram constructed with OA and OB as adjacent sides is **a**$\times$**b**, for it is twice the area of $\triangle$OAB.

2. The vector area of the triangle, the position vectors whose vertices are **a, b, c** is

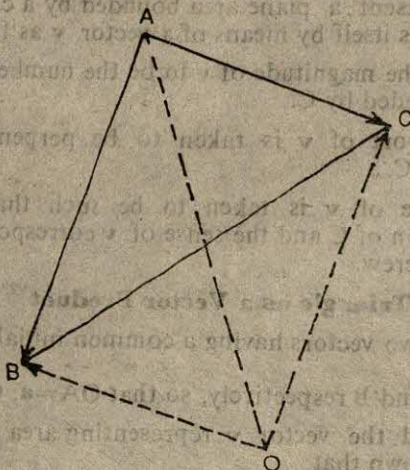


Fig. 8·37.

$$\tfrac{1}{2}\overrightarrow{AB}\times\overrightarrow{AC}$$

$$=\tfrac{1}{2}(\overrightarrow{OB}-\overrightarrow{OA})\times(\overrightarrow{OC}-\overrightarrow{OA})$$

$$=\tfrac{1}{2}(\mathbf{b}-\mathbf{a})\times(\mathbf{c}-\mathbf{a})$$

$$=\tfrac{1}{2}[\mathbf{b}\times\mathbf{c}+\mathbf{c}\times\mathbf{a}+\mathbf{a}\times\mathbf{b}].$$

Example 22. *Find the area of the parallelogram whose adjacent sides are determined by the vectors*

$$\mathbf{a}=\mathbf{i}+2\mathbf{j}+\mathbf{k},\ \mathbf{b}=-3\mathbf{i}-2\mathbf{j}+\mathbf{k}.$$

(*Roorkee Entrance, 1979*)

Solution.

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ 1 & 2 & 1\\ -3 & -2 & 1\end{vmatrix},$$

$$=4\mathbf{i}-4\mathbf{j}+4\mathbf{k}.$$

Required area $=|\mathbf{a}\times\mathbf{b}|$,

$$=\sqrt{(4^2+(-4)^2+4^2)}=4\sqrt{3}\text{ sq. units.}$$

8·10·6. Moment of a Force About a Point

A force is a vector quantity. It is completely specified by its vector and its point of application. It is important to note that in the case of a force it is not enough to specify only the magnitude, support and sense of the force, but we must also specify the point

of application of the force. Of course, by the principle of transmissibility of forces (which you must heve read in your study of Physics), it is enough to specify *any* one point on the line of action.

Let **F** be a given force and let P be any point on its line of action

Take any arbitrary point O and let $\vec{OP}=\mathbf{r}$.

The vector

$$\mathbf{M}=\mathbf{r}\times\mathbf{F}$$

is called the *moment vector* of the force **F** about the point O. The moment vector about O is independent of the choice of the point P on the line of action of F, *i.e.*, it will remain unaltered if we replace P by some other point, say Q, on the line of action of **F**. In fact, the moment vector in that case becomes

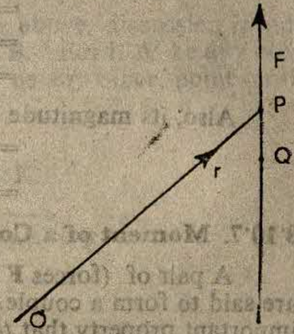


Fig. 8·38.

$$\vec{OQ}\times\mathbf{F}=(\vec{OP}+\vec{PQ})\times\mathbf{F},$$

$$=\vec{OP}\times\mathbf{F}+\vec{PQ}\times\mathbf{F},$$

$$=\mathbf{r}\times\mathbf{F},$$

since $\vec{PQ}$ and **F** are parallel vectors and consequently $\vec{PQ}\times\mathbf{F}=\mathbf{0}$.

Remarks 1. If **a** be the position vector of a point O and **b** that of any point on the line of application of **F**, then $\mathbf{r}=\mathbf{b}-\mathbf{a}$, so that $\mathbf{M}=(\mathbf{b}-\mathbf{a})\times\mathbf{F}$.

2. The moment vector is perpendicular to **F**.

3. The magnitude of the vector **M** is called the scalar moment of **F** about O.

4. A system of concurrent forces is equivalent to a single force whose point of application is the point of concurrence and whose vector is the sum of the given vectors.

The moment vector of such a system about a given point O can be obtained by taking the vector sum of the moment vectors of the individual forces about O.

Example 23. *Find the moment of the force* $F=3\mathbf{i}-2\mathbf{j}+4\mathbf{k}$ *acting at the point P whose position vector is* $-\mathbf{i}+\mathbf{j}-2\mathbf{k}$ *about the point O whose position vector is* $-2\mathbf{i}+5\mathbf{j}-\mathbf{k}$.

Solution. Here $\vec{OP}=(-\mathbf{i}+\mathbf{j}-2\mathbf{k})-(-2\mathbf{i}+5\mathbf{j}-\mathbf{k})$,

$$=\mathbf{i}-4\mathbf{j}-\mathbf{k},$$

$$\mathbf{F}=3\mathbf{i}-2\mathbf{j}+4\mathbf{k}.$$

Moment of **F** about O

$$=\vec{OP}\times\mathbf{F},$$
$$=(\mathbf{i}-4\mathbf{j}-\mathbf{k})\times(3\mathbf{i}-2\mathbf{j}+4\mathbf{k}),$$
$$=-18\mathbf{i}-7\mathbf{j}+10\mathbf{k}.$$

Also, its magnitude

$$=\sqrt{\{(-18)^2+(-7)^2+(10)^2\}},$$
$$=\sqrt{473}\,.$$

8·10·7. Moment of a Couple

A pair of (forces **F** and **—F** having different lines of action are said to form a couple. The moment sum of a couple has the important property that *the moment sum of the two forces* of a couple *is the same about every point.*

The above result can be proved easily.

Suppose the couple consists of two forces **F** and $-\mathbf{F}$. Let A be a point on the line of action of **F** and let B be a point on the line of application of $-\mathbf{F}$.

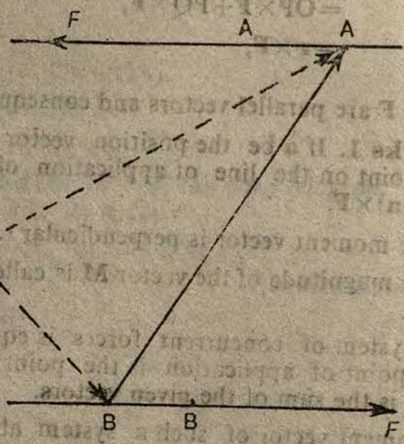


Fig. 8·39.

The moment of the couple about any point O.

$=$vector sum of the moments of the forces **F** and $-\mathbf{F}$ about O,

$$=\vec{OA}\times\mathbf{F}+\vec{OB}\times(-\mathbf{F}),$$
$$=(\vec{OA}-\vec{OB})\times\mathbf{F},$$
$$=\vec{BA}\times\mathbf{F}.$$

From (1) we find that the moment of the couple is independent of the choice of the point about which we take the moment.

Remark. The vector $\vec{BA}\times\mathbf{F}$ in the above discussion is independent of the choice of the points A and B. For if A′ be any other point on the line of action of **F**, and B′ be any other point on the line of action of $-\mathbf{F}$, then

$$\overrightarrow{B'A'}\times\mathbf{F}=(\overrightarrow{B'B}+\vec{BA}+\overrightarrow{AA'})\times\mathbf{F},$$

$$=\overrightarrow{B'B}\times\mathbf{F}+\vec{BA}\times\mathbf{F}+\overrightarrow{AA'}\times\mathbf{F},$$

$$=\vec{BA}\times\mathbf{F},$$

since $\overrightarrow{B'B}\times\mathbf{F}$ and $\overrightarrow{AA'}\times\mathbf{F}$ are both zero.

Example 24. *Find the moment of the couple consisting of the force* $\mathbf{F}=3\mathbf{i}+2\mathbf{j}-\mathbf{k}$ *acting through the point* $\mathbf{i}-\mathbf{j}+\mathbf{k}$ *and* $-\mathbf{F}$ *acting through the point* $2\mathbf{i}-3\mathbf{j}-\mathbf{k}$.

Solution. Let A be the point $\mathbf{i}-\mathbf{j}+\mathbf{k}$ and B be the point $2\mathbf{i}-3\mathbf{j}-\mathbf{k}$. The moment **M** of the couple

$$=\vec{BA}\times\mathbf{F},$$

$$=[2\mathbf{i}-3\mathbf{j}-\mathbf{k})-(\mathbf{i}-\mathbf{j}+\mathbf{k})]\times(3\mathbf{i}+2\mathbf{j}-\mathbf{k}),$$

$$=(\mathbf{i}-2\mathbf{j}-2\mathbf{k})\times(3\mathbf{i}+2\mathbf{j}-\mathbf{k}),$$

$$=6\mathbf{i}-5\mathbf{j}+8\mathbf{k}.$$

Also $|\mathbf{M}|=\sqrt{[6^2+(-5)^2+8^2]}$

$$=\sqrt{125}.$$

EXERCISE 8 (*f*)

1. Find the area of the triangle two of whose adjacent sides are determined by the vectors

   $\mathbf{i}+2\mathbf{j}-3\mathbf{k}$ and $3\mathbf{i}-4\mathbf{j}+5\mathbf{k}$.

2. Find the area of the parallelogram having diagonals

   $\mathbf{a}=3\mathbf{i}+\mathbf{j}-2\mathbf{k}$ and $\mathbf{b}=\mathbf{i}-3\mathbf{j}+4\mathbf{k}$. (*M.N.R. 1985*)

3. Find the area of the parallelogram whose adjacent sides are determined by the vectors

   $\mathbf{a}=\mathbf{i}-2\mathbf{j}+3\mathbf{k}$, $\mathbf{b}=-3\mathbf{i}-2\mathbf{j}+\mathbf{k}$. (*M.N.R. 1979*)

4. Prove that parallelograms on the same base and between the same parallels are equal in area.

5. In the sides BC, CA, AB of a triangle ABC, three points

D, E, F are taken such that each of BD, CE, AF is equal to one-third of the corresponding sides. Prove that

$$\triangle DEF = \frac{1}{3} \triangle ABC.$$

6. If D, E, F are mid-points of the sides BC, CA, AB of a triangle ABC, three points D, E, F are taken such that each of BD, CE, AF is equal to one-third of the corresponding side. Prove that

$$\triangle DEF = \frac{1}{3} \triangle ABC.$$

7. If AC and BD are two diagonals of a quadrilateral ABCD, show that its area is $\frac{1}{2}(\overrightarrow{AC} \times \overrightarrow{BD})$.

8. If $A_1, A_2, \ldots, A_n$ are the vertices of a regular plane polygon with n sides and O is the centre, then show that

$$\sum_{i=1}^{n-1} (\overrightarrow{OA_i} \times \overrightarrow{OA_{i+1}}) = (1-n)(\overrightarrow{OA_2} \times \overrightarrow{OA_1}).$$

(*I.I.T. J.E.E., 1982*)

9. E and F are the middle points of the sides AB and AC of a triangle ABC. CK is drawn parallel to AB and meets BF produced in K. Prove that

$$\triangle EFK = \triangle ECF = \tfrac{1}{4} \triangle ABC.$$

10. A force $\mathbf{F} = 2\mathbf{i} + \mathbf{j} - \mathbf{k}$ acts at a point A whose position vector is $2\mathbf{i} - \mathbf{j}$. Find the moment of $\mathbf{F}$ about the origin.

(*Roorkee Entrance, 1987*)

11. Find the moment of the force $\mathbf{P} = \mathbf{i} - \mathbf{j} + \mathbf{k}$ acting at a point A whose position vector is $4\mathbf{i} - 3\mathbf{j} - 2\mathbf{k}$ about the point (1, 0, 1).

(*Roorkee Entrance, 1980*)

12. Find the vector moment of three forces $\mathbf{i} + 2\mathbf{j} - 3\mathbf{k}$, $2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$, and $-\mathbf{i} - \mathbf{j} + \mathbf{k}$ acting on a particle at a point $\mathbf{j} + 2\mathbf{k}$, about the point $\mathbf{i} - 2\mathbf{j}$.

8·11. SCALAR TRIPLE PRODUCT

Let **a**, **b**, **c** be any three vectors. Consider the scalar product of the vectors $\mathbf{a} \times \mathbf{b}$ and **c**, *i.e.*, $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$. Two different cases arise, according as the vectors **a**, **b**, **c** are coplanar or non-coplanar.

Case I. Let **a**, **b**, **c** be coplanar. The vector $\mathbf{a} \times \mathbf{b}$ is perpendicular to the plane containing the vectors **a** and **b**, and is therefore perpendicular to **c** as well. Therefore $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = 0$.

Case II. Let **a**, **b**, **c** be non-coplanar. Let $\overrightarrow{OA}, \overrightarrow{OB}, \overrightarrow{OC}$ be three co-initial vectors equal to **a**, **b**, **c** respectively. Construct a parallelopiped with OA, OB, OC as three coterminus edges.

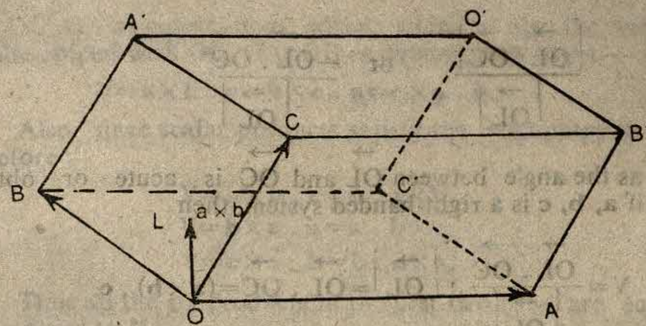


Fig. 8·40 (*a*)

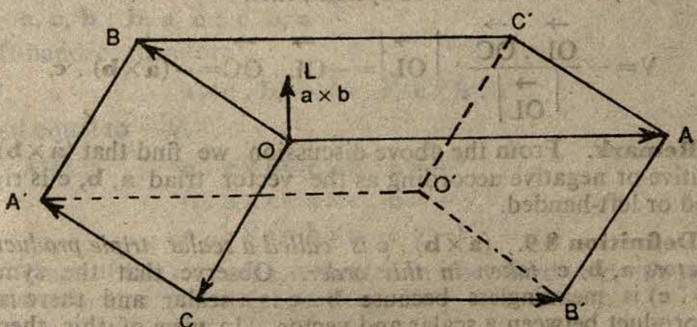


Fig. 8·40 (*b*)

Let V be the volume of the parallelopiped. We shall regard V as positive. We shall consider two different cases, according as the triad of vectors **a**, **b**, **c** is right-handed or left-handed. The vector $\mathbf{a}\times\mathbf{b}$ denotes the vector area of the parallelogram with OA and OB as adjacent sides.

The volume of the parallelopiped is by definition, the product of the area of this parallelogram and the perpendicular distance of C from the plane AOB.

Let $\qquad \vec{OL}=\mathbf{a}\times\mathbf{b}.$

The vector triad **a**, **b**, $\mathbf{a}\times\mathbf{b}$ is right-hand. The angle between the vectors $\mathbf{a}\times\mathbf{b}$ and **c** is acute or obtuse according as the vector triad **a**, **b**, **c** is right-handed [fig. 8·40(*a*)] or left-handed [fig. 8·40(*b*)]. The perpendicular distance of C from the plane AOB is equal to the absolute value of the projection of OC on OL, and therefore, it is equal to

$$\frac{\left|\vec{OL}\,.\,\vec{OC}\right|}{\left|\vec{OL}\right|}$$

i.e., $\dfrac{|\vec{OL}\,.\,\vec{OC}|}{|\vec{OL}|}$ or $\dfrac{-\vec{OL}\,.\,\vec{OC}}{|\vec{OL}|}$

according as the angle between $\vec{OL}$ and $\vec{OC}$ is acute or obtuse. Therefore if **a, b, c** is a right-handed system, then

$$V=\frac{\vec{OL}\,.\,\vec{OC}}{|\vec{OL}|}\cdot|\vec{OL}|=\vec{OL}\,.\,\vec{OC}=(\mathbf{a}\times\mathbf{b})\,.\,\mathbf{c}.$$

If **a, b, c** is a left-handed system, then

$$V=-\frac{\vec{OL}\,.\,\vec{OC}}{|\vec{OL}|}\cdot|\vec{OL}|=-\vec{OL}\,.\,\vec{OC}=-(\mathbf{a}\times\mathbf{b})\,.\,\mathbf{c}.$$

Remark. From the above discussion we find that $(\mathbf{a}\times\mathbf{b})\,.\,\mathbf{c}$ is positive or negative according as the vector triad **a, b, c** is right-handed or left-handed.

Definition 8·9. $(\mathbf{a}\times\mathbf{b})\,.\,\mathbf{c}$ *is called a scalar triple product of the vectors* **a, b, c** *taken in this order.* Observe that the symbol $\mathbf{a}\times(\mathbf{b}\,.\,\mathbf{c})$ is meaningless because $\mathbf{b}\,.\,\mathbf{c}$ is a scalar and there is no cross product between a scalar and vector. In view of this, there is no ambiguity if we simply write $\mathbf{a}\times\mathbf{b}\,.\,\mathbf{c}$ instead of $(\mathbf{a}\times\mathbf{b})\,.\,\mathbf{c}$. In future, we shall use the simpler notation $\mathbf{a}\times\mathbf{b}\,.\,\mathbf{c}$ instead of $(\mathbf{a}\times\mathbf{b})\,.\,\mathbf{c}$.

8·11·1. Different Scalar Triple Products Formed by Three Vectors

Given three different vectors **a, b, c** we can form the following twelve scalar triple products

$$\begin{array}{l}
\mathbf{a}\times\mathbf{b}\,.\,\mathbf{c},\ \mathbf{b}\times\mathbf{c}\,.\,\mathbf{a},\ \mathbf{c}\times\mathbf{a}\,.\,\mathbf{b},\\
\mathbf{a}\,.\,\mathbf{b}\times\mathbf{c},\ \mathbf{b}\,.\,\mathbf{c}\times\mathbf{a},\ \mathbf{c}\,.\,\mathbf{a}\times\mathbf{b},\\
\mathbf{a}\times\mathbf{c}\,.\,\mathbf{b},\ \mathbf{b}\times\mathbf{a}\,.\,\mathbf{c},\ \mathbf{c}\times\mathbf{b}\,.\,\mathbf{a},\\
\mathbf{a}\,.\,\mathbf{c}\times\mathbf{b},\ \mathbf{b}\,.\,\mathbf{a}\times\mathbf{c},\ \mathbf{c}\,.\,\mathbf{b}\times\mathbf{a}.
\end{array}$$

It turns out that all these products are not different. The six products written in the first two rows are all equal; also, the six products written in the last two rows are also equal, each being the negative of the products written in the first two rows.

In order to demonstrate the assertions made above, let us suppose that the vectors **a, b, c** form a right-handed triad. Then the vectors **b, c, a** and **c, a, b** also form right-handed triads.

Also, in the notation used above the parallelopiped with OA, OB, OC as coterminous edges is the same as the parallelopiped with

OB, OC, OA as coterminous edges, which is also the same as the parallelopiped with OC, OA, OB as coterminous edges.

$$\therefore \quad V = \mathbf{a} \times \mathbf{b} \cdot \mathbf{c} = \mathbf{b} \times \mathbf{c} \cdot \mathbf{a} = \mathbf{c} \times \mathbf{a} \cdot \mathbf{b}.$$

Also, since scalar products satisfy the commutative property, therefore

$$V = \mathbf{a} \times \mathbf{b} \cdot \mathbf{c} = \mathbf{c} \cdot \mathbf{a} \times \mathbf{b}$$
$$V = \mathbf{b} \times \mathbf{c} \cdot \mathbf{a} = \mathbf{a} \cdot \mathbf{b} \times \mathbf{c}$$
$$V = \mathbf{c} \times \mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{c} \times \mathbf{a}.$$

Thus all the six products in the first two rows are equal, each being equal to V.

Also since each of the triads

$\mathbf{a}, \mathbf{c}, \mathbf{b}$; $\mathbf{b}, \mathbf{a}, \mathbf{c}$; $\mathbf{c}, \mathbf{b}, \mathbf{a}$

is left-handed, therefore

$$\mathbf{a} \times \mathbf{c} \cdot \mathbf{b}, \ \mathbf{b} \times \mathbf{a} \cdot \mathbf{c}, \ \mathbf{c} \times \mathbf{b} \cdot \mathbf{a}$$

are all equal to $-V$.

Also
$$\mathbf{a} \cdot \mathbf{c} \times \mathbf{b} = \mathbf{c} \times \mathbf{b} \cdot \mathbf{a} = -V,$$
$$\mathbf{b} \cdot \mathbf{a} \times \mathbf{c} = \mathbf{a} \times \mathbf{c} \cdot \mathbf{b} = -V,$$
$$\mathbf{c} \cdot \mathbf{b} \times \mathbf{a} = \mathbf{b} \times \mathbf{a} \cdot \mathbf{c} = -V.$$

Thus all the six products written in the third and the fourth rows are equal, each being equal to $-V$.

In view of the above discussion we have the following two theorems :

Theorem 8·13. *In a scalar triple product the position of the dot and the cross can be interchanged without any change in the value of the scalar triple product.*

Theorem 8·14. *The value of a scalar triple product does not change when the vectors are permuted cyclically, and is changed in sign but remains unchanged numerically when the vectors are permuted anti-cyclically.*

Remarks 1. The scalar triple product $\mathbf{a} \times \mathbf{b} \cdot \mathbf{c}$ is usually denoted by $[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]$. This notation is justified because it takes note of the cyclic order of the vectors which is important and disregards the position of dot and cross which does not affect the value of the product. With this notation it is obvious that

$$[\mathbf{a}\ \mathbf{b}\ \mathbf{c}] = [\mathbf{b}\ \mathbf{c}\ \mathbf{a}] = [\mathbf{c}\ \mathbf{a}\ \mathbf{b}] = -[\mathbf{a}\ \mathbf{c}\ \mathbf{b}] = -[\mathbf{b}\ \mathbf{a}\ \mathbf{c}]$$
$$= -[\mathbf{c}\ \mathbf{b}\ \mathbf{a}].$$

2 The scalar triple product is zero if any two vectors are the same. For example,

$$[\mathbf{a}\ \mathbf{a}\ \mathbf{c}] = \mathbf{a} \times \mathbf{a} \cdot \mathbf{c} = 0,$$
$$[\mathbf{a}\ \mathbf{b}\ \mathbf{a}] = [\mathbf{a}\ \mathbf{a}\ \mathbf{b}] = 0.$$

3. The scalar triple product of three vectors is zero if and only if the vectors are coplanar.

8·11 2. Volume of a Parallelopiped

From the preceding discussion we have the following :

Theorem 8·15. *The volume V of the parallelopiped having three coterminous edges* **a, b, c** *is* $|(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}|$.

8·11·3. An Expression for the Scalar Triple Product of Vectors in Terms of Components

Let **i, j, k** be an orthonormal triad of vectors. Suppose three vectors **a, b, c** are expressible as

$$\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k},$$
$$\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k},$$
$$\mathbf{c}=c_1\mathbf{i}+c_2\mathbf{j}+c_3\mathbf{k}.$$

Then $\mathbf{b}\times\mathbf{c}=(b_2c_3-b_3c_2)\,\mathbf{i}+(b_3c_1-b_1c_3)\,\mathbf{j}+(b_1c_2-b_2c_1)\,\mathbf{k},$

so that $\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}=a_1(b_2c_3-b_3c_2)+a_2(b_3c_1-b_1c_3)+a_3(b_1c_2-b_2c_1)$

$$=\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

Example 25. *If* $\mathbf{a}=3\mathbf{i}-\mathbf{j}+2\mathbf{k}$, $\mathbf{b}=2\mathbf{i}-\mathbf{j}+3\mathbf{k}$, $\mathbf{c}=-\mathbf{i}-2\mathbf{j}+\mathbf{k}$, *find* [**a b c**].

Solution.

$$[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]=\begin{vmatrix} 3 & -1 & 2 \\ 2 & -1 & 3 \\ -1 & -2 & 1 \end{vmatrix}$$

$$=3(-1\cdot 1-3\cdot(-2))$$
$$-(-1)(2\cdot 1-3(-1))$$
$$+2(2(-2)-(-1)(-1))$$
$$=15+5-10$$
$$=10.$$

Example 26. *Find the volume of the parallelopiped having* $\overrightarrow{OA}=2\mathbf{i}-3\mathbf{j}$, $\overrightarrow{OB}=\mathbf{i}+\mathbf{j}-\mathbf{k}$, $\overrightarrow{OC}=3\mathbf{i}-\mathbf{k}$ *as three edges.*

(*I.I.T. J.E.E., 1983*)

Solution. The volume of the parallelopiped having $\overrightarrow{OA}$, $\overrightarrow{OB}$, $\overrightarrow{OC}$ as three coterminous edges

$$=\left[\vec{OA}\ \vec{OB}\ \vec{OC}\right]$$

$$=\begin{vmatrix} 2 & -3 & 0 \\ 1 & 1 & -1 \\ 3 & 0 & -1 \end{vmatrix}$$

$$=2\,(1(-1)-0\,.\,(-1))+3\,(1(-1)-(-1)3)$$

$$=4.$$

Example 27. *Show that the four points whose position vectors are* $3\mathbf{i}+9\mathbf{j}+4\mathbf{k}$, $-(\mathbf{j}+\mathbf{k})$, $4(-\mathbf{i}+\mathbf{j}+\mathbf{k})$ *and* $4\mathbf{i}+5\mathbf{j}+\mathbf{k}$ *are coplanar.*

Solution. Let us name the four points A, B, C and D respectively. If O be the origin of reference, then

$$\vec{OA}=3\mathbf{i}+9\mathbf{j}+4\mathbf{k},$$

$$\vec{OB}=-\mathbf{j}-\mathbf{k}$$

$$\vec{OC}=-4\mathbf{i}+4\mathbf{j}+4\mathbf{k}$$

$$\vec{OD}=4\mathbf{i}+5\mathbf{j}+\mathbf{k}$$

$$\therefore \quad \vec{AB}=-3\mathbf{i}-10\mathbf{j}-5\mathbf{k},$$

$$\vec{AC}=-7\mathbf{i}-5\mathbf{j}$$

$$\vec{AD}=\mathbf{i}-4\mathbf{j}-3\mathbf{k}.$$

$$\vec{AB}\,.\,(\vec{AC}\times\vec{AD})$$

$$=\begin{vmatrix} -3 & -10 & -5 \\ -7 & -5 & 0 \\ 1 & -4 & -3 \end{vmatrix}$$

$$=-3\,((-5)(-3)-0\,.\,(-4))$$
$$+7\,((-10)\,.\,(-3)-(-5)(-4))$$
$$+1\,((-10)\,.\,0-(-5)(-5))$$

$$=-45+70-25$$

$$=0.$$

Since $\vec{AB}\,.\,(\vec{AC}\times\vec{AD})=0$, therefore the lines AB, AC, AD are coplanar, and consequently the points A, B, C, D are coplanar.

Example 28. *Prove that*

$$(\mathbf{a}+\mathbf{b}) \cdot (\mathbf{b}+\mathbf{c})\times(\mathbf{c}+\mathbf{a})=2\,[\mathbf{a}\ \mathbf{b}\ \mathbf{c}].$$

Solution.

$$\begin{aligned}\text{L.H.S.} &= (\mathbf{a}+\mathbf{b}) \cdot \{(\mathbf{b}+\mathbf{c})\times(\mathbf{c}+\mathbf{a})\}\\ &= (\mathbf{a}+\mathbf{b}) \cdot \{\mathbf{b}\times\mathbf{c}+\mathbf{b}\times\mathbf{a}+\mathbf{c}\times\mathbf{a}\}, \text{ since } \mathbf{c}\times\mathbf{c}=0\\ &= \mathbf{a} \cdot (\mathbf{b}\times\mathbf{c})+\mathbf{b} \cdot (\mathbf{b}\times\mathbf{c})+\mathbf{a} \cdot (\mathbf{b}\times\mathbf{a})\\ &\quad +\mathbf{b} \cdot (\mathbf{b}\times\mathbf{a})+\mathbf{a} \cdot (\mathbf{c}\times\mathbf{a})+\mathbf{b} \cdot (\mathbf{c}\times\mathbf{a}),\\ &= \mathbf{a} \cdot (\mathbf{b}\times\mathbf{c})+\mathbf{b} \cdot (\mathbf{c}\times\mathbf{a}), \text{ since the other terms all vanish}\\ &= [\mathbf{a}\ \mathbf{b}\ \mathbf{c}]+[\mathbf{b}\ \mathbf{c}\ \mathbf{a}]\\ &= 2\,[\mathbf{a}\ \mathbf{b}\ \mathbf{c}], \text{ since } [\mathbf{b}\ \mathbf{c}\ \mathbf{a}]=[\mathbf{a}\ \mathbf{b}\ \mathbf{c}].\end{aligned}$$

EXERCISE 8 (*g*)

1. Compute the following scalar triple products :
   (*a*) $(3\mathbf{i}+\mathbf{j}-2\mathbf{k})\times(-\mathbf{i}+2\mathbf{j}+\mathbf{k}) \cdot (\mathbf{i}+\mathbf{j})$.
   (*b*) $(2\mathbf{i}+\mathbf{j}+\mathbf{k}) \cdot (2\mathbf{i}-3\mathbf{j}+\mathbf{k})\times(\mathbf{i}-\mathbf{j}+2\mathbf{k})$.
2. If $\mathbf{a}=-2\mathbf{i}+3\mathbf{j}-4\mathbf{k}$, $\mathbf{b}=-\mathbf{j}+2\mathbf{k}$, $\mathbf{c}=\mathbf{i}-2\mathbf{j}+3\mathbf{k}$, find $[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]$.
3. Show that each of the following triads of vectors **a**, **b**, **c** is coplanar :
   (*a*) $\mathbf{a}=\mathbf{i}-3\mathbf{j}+4\mathbf{k}$, $\mathbf{b}=2\mathbf{i}-\mathbf{j}+2\mathbf{k}$, $\mathbf{c}=4\mathbf{i}-7\mathbf{j}+10\mathbf{k}$.
   (*b*) $\mathbf{a}=2\mathbf{i}+3\mathbf{j}-\mathbf{k}$, $\mathbf{b}=-3\mathbf{i}+\mathbf{j}-4\mathbf{k}$, $\mathbf{c}=-\mathbf{i}+4\mathbf{j}-5\mathbf{k}$.
4. Find the volume of the parallelopiped having **a**, **b**, **c** as coterminous edges, when
   (*a*) $\mathbf{a}=2\mathbf{i}+\mathbf{j}-\mathbf{k}$, $\mathbf{b}=\mathbf{i}+2\mathbf{j}+3\mathbf{k}$, $\mathbf{c}=-3\mathbf{i}-\mathbf{j}+\mathbf{k}$.
   (*b*) $\mathbf{a}=4\mathbf{i}-\mathbf{j}+\mathbf{k}$, $\mathbf{b}=2\mathbf{i}+\mathbf{j}-3\mathbf{k}$, $\mathbf{c}=2\mathbf{i}-2\mathbf{j}+6\mathbf{k}$.
5. The position vectors of points O, A, B and C are $\mathbf{i}+\mathbf{j}$, $2\mathbf{i}-3\mathbf{j}+5\mathbf{k}$, $3\mathbf{i}+\mathbf{j}-4\mathbf{k}$, and $6\mathbf{i}+3\mathbf{j}-\mathbf{k}$ respectively. Find the volume of the parallelopiped having OA, OB, OC as three coterminous edges.

8·12. VECTOR TRIPLE PRODUCT

If **a**, **b**, **c** are any three vectors, then the symbols $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ is meaningful because $\mathbf{b}\times\mathbf{c}$ is a vector and $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ is also a vector. Similarly $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ is also meaningful. These products are called **vector triple products** because they arise by taking vector products of cetain vectors obtained from three vectors. These products can be expressed as linear combinations of vectors **a**, **b** and **c**. We shall try to see as to how this can be done.

Let us introduce an orthonormal basis **i**, **j**, **k** as follows :

(1) Take a unit vector along **a** and call it **i**.

(2) Take a unit vector perpendicular to **a** lying in the plane determined by **a** and **b**, and call it **j**.

(3) Take $\mathbf{k}$ to be the unit vector perpendicular to $\mathbf{i}$ and $\mathbf{j}$, such $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ form a right-handed triad.

With the above notation, we have

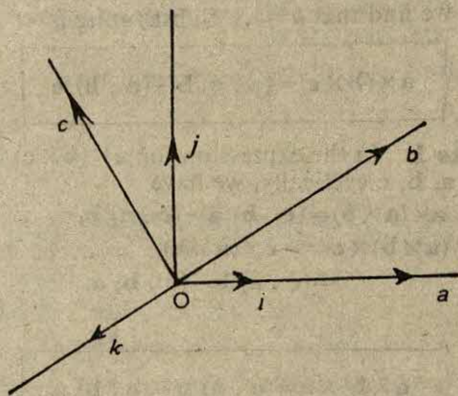


Fig. 8·41.

$$\mathbf{a}=a_1\mathbf{i}.$$
$$\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j},$$
$$\mathbf{c}=c_1\mathbf{i}+c_2\mathbf{j}+c_3\mathbf{k}$$

Now

$$\mathbf{b}\times\mathbf{c}=(b_1\mathbf{i}+\mathbf{b}_2\mathbf{j})\times(c_1\mathbf{i}+c_2\mathbf{j}+c_3\mathbf{k})$$
$$=b_2c_3\mathbf{i}-b_1c_3\mathbf{j}+(b_1c_2-b_2c_1)\,\mathbf{k}$$

$$\therefore\quad \mathbf{a}\times(\mathbf{b}\times\mathbf{c})=a_1\mathbf{i}\times[b_2c_3\mathbf{i}-b_1c_3\mathbf{j}+(b_1c_2-b_2c_1)\,\mathbf{k}]$$
$$=-a_1(b_1c_2-b_2c_1)\,\mathbf{j}-a_1b_1c_3\,\mathbf{k}. \qquad ...(1)$$

Now $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ is perpendicular to $\mathbf{a}$, and therefore the expression for $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ as a linear combination of $\mathbf{b}$ and $\mathbf{c}$ must be of the form $p\mathbf{b}+q\mathbf{c}$ for some scalars p and q. We shall try to find the values of p and q.

Now

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=p\mathbf{b}+q\mathbf{c} \qquad ...(2)$$

Taking scalar product of both sides of (2) with $\mathbf{a}$ we have

$$\mathbf{a}\,.\,[\mathbf{a}\times(\mathbf{b}\times\mathbf{c})]=\mathbf{a}\,.\,(p\mathbf{b}+q\mathbf{c}),$$

or $\quad 0=p\,(\mathbf{a}\,.\,\mathbf{b})+q\,(\mathbf{a}\,.\,\mathbf{c}),$

which gives $\quad p=(\mathbf{a}\,.\,\mathbf{c})h, \qquad q=-(\mathbf{b}\,.\,\mathbf{c})h,$

for some scalar h. ...(3)

From (2) and (3) we have

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=h\,[(\mathbf{a}\,.\,\mathbf{c})\,\mathbf{b}-(\mathbf{a}\,.\,\mathbf{b})\,\mathbf{c}], \qquad ...(4)$$

where h is to be determined.

Now $\quad (\mathbf{a}\,.\,\mathbf{c})=a_1c_1,\ \mathbf{a}\,.\,\mathbf{b}=a_1b_1,$

so that $(\mathbf{a}\,.\,\mathbf{c})\,\mathbf{b}-(\mathbf{a}\,.\,\mathbf{b})\,\mathbf{c}=a_1c_1(b_1\mathbf{i}+b_2\mathbf{j})-a_1b_1(c_1\mathbf{i}+c_2\mathbf{j}+c_3\mathbf{k})$

$$=-a_1(b_1c_2\mid b_2c_1)\,\mathbf{j}-a_1b_1c_3\mathbf{k} \qquad ...(5)$$

From (4) and (5), we have

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=h\,[-a_1(b_1c_2-b_2c_1)\,\mathbf{j}-a_1b_1c_3\,\mathbf{k}] \qquad \ldots(6)$$

Comparing the two expressions for $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ obtained in (1) and (6) above we find that $h=1$. Substituting $h=1$ in (4), we have

$$\boxed{\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=(\mathbf{a}\,.\,\mathbf{c})\,\mathbf{b}-(\mathbf{a}\,.\,\mathbf{b})\,\mathbf{c}} \qquad \ldots(A)$$

Remarks 1. In the expression for $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ obtained above, by permuting **a, b, c** cyclically, we have

$$\mathbf{c}\times(\mathbf{a}\times\mathbf{b})=(\mathbf{c}\,.\,\mathbf{b})\,\mathbf{a}-(\mathbf{c}\,.\,\mathbf{a})\,\mathbf{b},$$

so that

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\times\mathbf{c}&=-\mathbf{c}\times(\mathbf{a}\times\mathbf{b})\\&=(\mathbf{c}\,.\,\mathbf{a})\,\mathbf{b}-(\mathbf{c}\,.\,\mathbf{b})\,\mathbf{a}.\end{aligned}$$

Thus

$$\boxed{(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=(\mathbf{c}\,.\,\mathbf{a})\,\mathbf{b}-(\mathbf{c}\,.\,\mathbf{b})\,\mathbf{a}} \qquad \ldots(B)$$

2. Comparing (A) and (B), we have

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})-(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=(\mathbf{c}\,.\,\mathbf{b})\,\mathbf{a}-(\mathbf{a}\,.\,\mathbf{b})\,\mathbf{c}, \qquad \ldots(C)$$

which shows that in general

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\neq(\mathbf{a}\times\mathbf{b})\times\mathbf{c},$$

i.e., vector multiplication is not, in general, associative.

3. From (C) above we find that if the right hand side is zero then **a** must be parallel to **c**, *i.e.*, $\mathbf{c}=m\mathbf{a}$ for some scalar m. Further more, if

$$\begin{aligned}\mathbf{c}=m\mathbf{a},\ \text{then}\quad &(\mathbf{c}\,.\,\mathbf{a})\,\mathbf{a}-(\mathbf{a}\,.\,\mathbf{b})\,\mathbf{c}\\&=(m\mathbf{a}\,.\,\mathbf{b})\,\mathbf{a}-(\mathbf{a}\,.\,\mathbf{b})\,(m\mathbf{a})\\&=m\,[(\mathbf{a}\,.\,\mathbf{b})\,\mathbf{a}-(\mathbf{a}\,.\,\mathbf{b})\,\mathbf{a}]\\&=m\,\mathbf{0}\\&=\mathbf{0}.\end{aligned}$$

Thus, $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=(\mathbf{a}\times)\mathbf{b}\times\mathbf{c}$ if and only if **a** and **c** are parallel.

8·12·1. Distributivity of Vector Multiplication over Addition

We have already seen that vector multiplication distributes itself over addition. We shall now give an alternative proof of this result by using the notion of scalar triple products.

Let **a, b, c** be three given vectors. We shall show that

$$\mathbf{a}\times(\mathbf{b}+\mathbf{c})=\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}.$$

Let us write

$$\mathbf{p}=\mathbf{a}\times(\mathbf{b}+\mathbf{c})-\mathbf{a}\times\mathbf{b}-\mathbf{a}\times\mathbf{c}.$$

We wish to show that $\mathbf{p}=\mathbf{0}$. Let **r** be *any* vector whatsoever.

Then $\mathbf{r} \cdot \mathbf{p}=\mathbf{r} \cdot [\mathbf{a}\times(\mathbf{b}+\mathbf{c})-\mathbf{a}\times\mathbf{b}-\mathbf{a}\times\mathbf{c}]$

$=\mathbf{r} \cdot \mathbf{a}\times(\mathbf{b}+\mathbf{c})-\mathbf{r} \cdot \mathbf{a}\times\mathbf{b}-\mathbf{r} \cdot \mathbf{a}\times\mathbf{c}$, by distributivity of dot product over addition

$=\mathbf{r}\times\mathbf{a} \cdot (\mathbf{b}+\mathbf{c})-\mathbf{r}\times\mathbf{a} \cdot \mathbf{b}-\mathbf{r}\times\mathbf{a} \cdot \mathbf{c}$, by interchanging the positions of cross and dot

$=\mathbf{r}\times\mathbf{a} \cdot [\mathbf{b}+\mathbf{c}-\mathbf{b}-\mathbf{c}]$, by distributivity of dot product

$=\mathbf{r}\times\mathbf{a} \cdot \mathbf{0}$

$=0.$

Since $\mathbf{r} \cdot \mathbf{p}=0$ for every vector $\mathbf{r}$ therefore $\mathbf{p}$ must be the zero vector. Since $\mathbf{p}=\mathbf{0}$, it follows that vector multiplication distributes itself over addition.

Example 29. *Solve the equation*

$$\mathbf{r}\times\mathbf{b}=\mathbf{a}$$

where the vectors **a**, **b** *are perpendicular to each other.*

Solution. The vector $\mathbf{a}\times\mathbf{b}$ is perpendicular to both **a** as well as **b**. Since **a** and **b** are mutually orthogonal, therefore $\mathbf{a}, \mathbf{b}, \mathbf{a}\times\mathbf{b}$ form a mutually orthogonal triad of vectors, and consequently every vector can be expressed as a linear combination of these three vectors.

Let $\quad \mathbf{r}=l\,\mathbf{a}+m\mathbf{b}+n(\mathbf{a}\times\mathbf{b}).$ ...(1)

Substituting the above value of $\mathbf{r}$ in the given equation, we have

$$[l\mathbf{a}+m\mathbf{b}+n(\mathbf{a}\times\mathbf{b})]\times\mathbf{b}=\mathbf{a}$$

$\Rightarrow \quad l(\mathbf{a}\times\mathbf{b})+n(\mathbf{a}\times\mathbf{b})\times\mathbf{b}=\mathbf{a}$

$\Rightarrow \quad l(\mathbf{a}\times\mathbf{b})+n[\mathbf{b}.\mathbf{b})\mathbf{a}-(\mathbf{b} \cdot \mathbf{a})\,\mathbf{b}]=\mathbf{a},$

$\Rightarrow \quad [n(\mathbf{b} \cdot \mathbf{b})+l]\,\mathbf{a}+l(\mathbf{a}\times\mathbf{b})=\mathbf{0},$

since $\quad \mathbf{a} \cdot \mathbf{b}=0.$

Since the vectors $\mathbf{a}, \mathbf{b}, \mathbf{a}\times\mathbf{b}$ are linearly independent, therefore the coefficients of $\mathbf{a}, \mathbf{b}, \mathbf{a}\times\mathbf{b}$ vanish separately.

Therefore

$$n(\mathbf{b} \cdot \mathbf{b})-1=0,\ l=0.$$

Thus $\quad l=0,\ n=(\mathbf{b} \cdot \mathbf{b})^{-1}.$ ...(2)

Observe that m does not appear in the above relations and therefore remains undetermined. Substituting the values of l and n in (1) we have

$$\mathbf{r}=m\mathbf{b}+(\mathbf{b} \cdot \mathbf{b})^{-1}\,(\mathbf{a}\times\mathbf{b}),$$

as the desired solution, where m is a parameter.

Example 30. *Show that for every triad of vectors* **a**, **b**, **c**,

$$[\mathbf{b}\times\mathbf{c}\ \ \mathbf{c}\times\mathbf{a}\ \ \mathbf{a}\times\mathbf{b}]=[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]^2.$$

Solution.

Write $\mathbf{b}\times\mathbf{c}=\mathbf{p},\ \mathbf{c}\times\mathbf{a}=\mathbf{q},\ \mathbf{a}\times\mathbf{b}=\mathbf{r}$

$$\begin{aligned}\mathbf{p}\times\mathbf{q}&=\mathbf{p}\times(\mathbf{c}\times\mathbf{a})\\&=\mathbf{p}\,.\,\mathbf{a}\ \mathbf{c}-\mathbf{p}\,.\,\mathbf{c}\ \mathbf{a}\\&=(\mathbf{b}\times\mathbf{c})\,.\,\mathbf{a}\ \mathbf{c}-(\mathbf{b}\times\mathbf{c})\,.\,\mathbf{c}\ \mathbf{a}\\&=[\mathbf{b}\ \mathbf{c}\ \mathbf{a}]\ \mathbf{b},\ \text{since } \mathbf{b}\times\mathbf{c} \text{ is perpendicular to } \mathbf{c}\end{aligned}$$

Now, $(\mathbf{p}\times\mathbf{q})\,.\,\mathbf{r}=[\mathbf{b}\ \mathbf{c}\ \mathbf{a}]\ \mathbf{c}\,.\,(\mathbf{a}\times\mathbf{b})$

$$\begin{aligned}&=[\mathbf{b}\ \mathbf{c}\ \mathbf{a}][\mathbf{c}\ \mathbf{a}\ \mathbf{b}]\\&=[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]^2,\end{aligned}$$

since $[\mathbf{b}\ \mathbf{c}\ \mathbf{a}]=[\mathbf{c}\ \mathbf{a}\ \mathbf{b}]=[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]$.

Example 31. *Prove that*

$$(\mathbf{a}\times\mathbf{b})\,.\,(\mathbf{c}\times\mathbf{d})=\begin{vmatrix}\mathbf{a}\,.\,\mathbf{c} & \mathbf{a}\,.\,\mathbf{d}\\ \mathbf{b}\,.\,\mathbf{c} & \mathbf{b}\,.\,\mathbf{d}\end{vmatrix}$$

Solution. Let us denote the vector $\mathbf{c}\times\mathbf{d}$ by $\mathbf{p}$. We then have

$$\begin{aligned}\text{L.H.S.}&=(\mathbf{a}\times\mathbf{b})\,.\,\mathbf{p}\\&=\mathbf{a}\,.\,(\mathbf{b}\times\mathbf{p}),\ \text{interchanging the positions of dot and cross}\\&=\mathbf{a}\,.\,[\mathbf{b}\times(\mathbf{c}\times\mathbf{d})]\\&=\mathbf{a}\,.\,[(\mathbf{b}\,.\,\mathbf{d})\mathbf{c}-(\mathbf{b}\,.\,\mathbf{c})\mathbf{d}]\\&=(\mathbf{b}\,.\,\mathbf{d})(\mathbf{a}\,.\,\mathbf{c})-(\mathbf{b}\,.\,\mathbf{c})(\mathbf{a}\,.\,\mathbf{d})\\&=\text{R.H.S.}\end{aligned}$$

Remark. The identity in the above example is due to the French mathematician Joseph Louis Lagrange, and is often called Lagrange's identity.

Example 32. *Show that*

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d})&=[\mathbf{a}\ \mathbf{b}\ \mathbf{d}]\mathbf{c}-[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]\mathbf{d}\\&=[\mathbf{a}\ \mathbf{c}\ \mathbf{d}]\mathbf{b}-[\mathbf{b}\ \mathbf{c}\ \mathbf{d}]\mathbf{a},\end{aligned}$$

for any four vectors **a, b, c, d.**

Solution. Writing $\mathbf{a}\times\mathbf{b}=\mathbf{p}$, we have

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d})&=\mathbf{p}\times(\mathbf{c}\times\mathbf{d})\\&=\mathbf{p}\,.\,\mathbf{d}\mathbf{c}-\mathbf{p}\,.\,\mathbf{c}\mathbf{d}.\\&=[\mathbf{a}\ \mathbf{b}\ \mathbf{d}]\mathbf{c}-[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]\mathbf{d}\qquad\cdots(1)\end{aligned}$$

Again, writing $\mathbf{c}\times\mathbf{d}=\mathbf{q}$, we have

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d})&=(\mathbf{a}\times\mathbf{b})\times\mathbf{q}\\&=\mathbf{q}\,.\,\mathbf{a}\ \mathbf{b}-\mathbf{q}\,.\,\mathbf{b}\ \mathbf{a}\\&=\mathbf{a}\,.\,\mathbf{q}\ \mathbf{b}-\mathbf{b}\,.\,\mathbf{q}\ \mathbf{a}\\&=[\mathbf{a}\ \mathbf{c}\ \mathbf{d}]\ \mathbf{b}-[\mathbf{b}\ \mathbf{c}\ \mathbf{d}]\mathbf{a}\qquad\cdots(2)\end{aligned}$$

Remark. Equating the two expressions for $(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d})$ obtained above, we have the identity,

$$[\mathbf{a}\ \mathbf{b}\ \mathbf{d}]\mathbf{c}-[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]\mathbf{d}=[\mathbf{a}\ \mathbf{c}\ \mathbf{d}]\mathbf{b}-[\mathbf{b}\ \mathbf{c}\ \mathbf{d}]\mathbf{a}$$

or $$[\mathbf{b}\ \mathbf{c}\ \mathbf{d}]\mathbf{a}-[\mathbf{a}\ \mathbf{c}\ \mathbf{d}]\mathbf{b}+[\mathbf{a}\ \mathbf{b}\ \mathbf{d}]\mathbf{c}-[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]\ \mathbf{d}=\mathbf{0}.$$

EXERCISE 8 (*h*)

1. Compute the following vector triple products :

   (*a*) $[(2\mathbf{i}+3\mathbf{j})\times(3\mathbf{j}-\mathbf{k})]\times(\mathbf{i}+\mathbf{k})$.

   (*b*) $[(\mathbf{i}-3\mathbf{j})\times(\mathbf{j}+\mathbf{k})]\times(\mathbf{i}-2\mathbf{j}+3\mathbf{k})$.

2. Given that $\mathbf{a}=2\mathbf{i}+\mathbf{j}-3\mathbf{k}$, $\mathbf{b}=\mathbf{i}-\mathbf{j}+2\mathbf{k}$, $\mathbf{c}=3\mathbf{i}+2\mathbf{j}-\mathbf{k}$, calculate the following :

   (*a*) $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ (*b*) $\mathbf{b}\times(\mathbf{c}\times\mathbf{a})$ (*c*) $\mathbf{c}\times(\mathbf{a}\times\mathbf{b})$

   (*d*) $(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$ (*e*) $(\mathbf{b}\times\mathbf{c})\times\mathbf{a}$ (*f*) $(\mathbf{c}\times\mathbf{a})\times\mathbf{b}$.

3. If $\mathbf{a}=3\mathbf{i}-\mathbf{j}+2\mathbf{k}$, $\mathbf{b}=2\mathbf{i}+\mathbf{j}-\mathbf{k}$, $\mathbf{c}=\mathbf{i}-2\mathbf{j}+2\mathbf{k}$, show by actual calculation that $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\neq(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$.

4. If $\mathbf{a}=2\mathbf{i}-5\mathbf{j}+7\mathbf{k}$, $\mathbf{b}=\mathbf{i}+\mathbf{j}-3\mathbf{k}$, $\mathbf{c}=4\mathbf{i}-10\mathbf{j}+14\mathbf{k}$, verify that $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$.

5. Show that the vectors $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$, $\mathbf{b}\times(\mathbf{c}\times\mathbf{a})$, $\mathbf{c}\times(\mathbf{a}\times\mathbf{b})$ are coplanar.

   Prove the following :

6. $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})+\mathbf{b}\times(\mathbf{c}\times\mathbf{a})+\mathbf{c}\times(\mathbf{a}\times\mathbf{b})=\mathbf{0}$.

7. $(\mathbf{b}\times\mathbf{c})\times(\mathbf{c}\times\mathbf{a})=[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]\mathbf{c}$.

8. $\mathbf{a}\times[\mathbf{b}\times(\mathbf{c}\times\mathbf{a})]=(\mathbf{a}\ .\ \mathbf{b})(\mathbf{a}\times\mathbf{c})$.

9. $[\mathbf{a}\times\mathbf{b}\ \ \mathbf{a}\times\mathbf{c}\ \ \mathbf{d}]=(\mathbf{a}\ .\ \mathbf{d})[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]$.

10. $[\mathbf{a}\times\mathbf{a}\times(\mathbf{a}\times\mathbf{b})]=(\mathbf{a}\ .\ \mathbf{a})\ (\mathbf{b}\times\mathbf{a})$.

11. $\mathbf{a}\times\{\mathbf{b}\times(\mathbf{c}\times\mathbf{d})\}=(\mathbf{b}\ .\ \mathbf{d})(\mathbf{a}\times\mathbf{c})-(\mathbf{b}\ .\ \mathbf{c})\ (\mathbf{a}\times\mathbf{d})$.

12. If $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ be three unit vectors such that
    $$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=\tfrac{1}{2}\mathbf{b}\ ,$$
    find the angle which $\mathbf{a}$ makes with $\mathbf{b}$ and $\mathbf{c}$, $\mathbf{b}$ and $\mathbf{c}$ being non-parallel.

13. If the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ are coplanar, show that
    $$(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d})=\mathbf{0}.$$

TEST YOUR UNDERSTANDING VIII

In each of the following problems four alternatives are given, out of which only one is correct. Put a tick mark (✓) against the correct alternative.

1. The position vectors of a pair of opposite vertices of a parallelogram are $2\mathbf{a}-3\mathbf{b}+\mathbf{c}$ and $-4\mathbf{a}+3\mathbf{b}+7\mathbf{c}$. The position vector of the mid-point of the diagonal determined by the other two vertices of the parallelogram is

(*a*) $-2\mathbf{a}+8\mathbf{c}$ (*b*) $6\mathbf{a}-6\mathbf{b}-6\mathbf{c}$
(*c*) $3\mathbf{a}-3\mathbf{b}-3\mathbf{c}$ (*d*) $-\mathbf{a}+4\mathbf{c}$.

2. The position vectors of A and B are $\mathbf{i}-\mathbf{j}+8\mathbf{k}$ and $-2\mathbf{i}+3\mathbf{j}-4\mathbf{k}$ respectively. The magnitude of $\overrightarrow{AB}$ is
(*a*) 15 (*b*) 19
(*c*) 11 (*d*) 13.

3. The vectors $\mathbf{a}$ and $\mathbf{b}$ include an angle of 60°. If $|\mathbf{a}|=2$, $|\mathbf{b}|=\sqrt{3}$, the value of $\mathbf{a}\cdot\mathbf{b}$ is
(*a*) 3 (*b*) 2
(*c*) $\sqrt{3}$ (*d*) $2\sqrt{3}$.

4. If $\mathbf{i}, \mathbf{j}, \mathbf{k}$ be an orthonormal triad of vectors, the value of $\mathbf{i}\times(\mathbf{j}\times\mathbf{k})+\mathbf{j}\times(\mathbf{k}\times\mathbf{i})+\mathbf{k}\times(\mathbf{i}\times\mathbf{j})$ is
(*a*) 0 (*b*) 1
(*c*) 3 (*d*) -1.

5. The vectors $2\mathbf{i}-3\mathbf{j}+4\mathbf{k}$ and $4\mathbf{i}-x\mathbf{j}+\mathbf{k}$ are perpendicular to each other. The value of x is
(*a*) 4 (*b*) -4
(*c*) 0 (*d*) -3.

6. The vectors $x\mathbf{i}-6\mathbf{j}+8\mathbf{k}$ and $2\mathbf{i}+3\mathbf{j}-4\mathbf{k}$ are parallel to each other. The value of x is
(*a*) -4 (*b*) 4
(*c*) 3 (*d*) 6.

7. The volume of the parallelopiped having $2\mathbf{i}-3\mathbf{j}+4\mathbf{k}$, $4\mathbf{i}+\mathbf{j}-5\mathbf{k}$ and $\mathbf{i}-4\mathbf{j}+3\mathbf{k}$ as three coterminous edges is
(*a*) 100 (*b*) 78
(*c*) 30 (*d*) -30.

8. The vectors $9\mathbf{i}-x\mathbf{j}+3\mathbf{k}$, $\mathbf{i}+\mathbf{j}-\mathbf{k}$, and $2\mathbf{i}-\mathbf{j}+4\mathbf{k}$ are coplanar. The value of x is
(*a*) 3 (*b*) 2
(*c*) 3/2 (*d*) $-3/2$.

9. The diagonals of a parallelogram are given by the vectors $\mathbf{a}=2\mathbf{i}+3\mathbf{j}-\mathbf{k}$, $\mathbf{b}=4\mathbf{i}-\mathbf{j}+2\mathbf{k}$. The area of the parallelogram is
(*a*) 16 (*b*) $\sqrt{285}$
(*c*) 17 (*d*) $\frac{1}{2}\sqrt{285}$

10. Two sides of a triangle are given by the vectors $\mathbf{i}-\mathbf{j}+\mathbf{k}$, $2\mathbf{i}-3\mathbf{j}+4\mathbf{k}$. Its area is
(*a*) 6 (*b*) $\frac{1}{2}\sqrt{6}$
(*c*) $\sqrt{6}$ (*d*) $2\sqrt{6}$.

REVIEW EXERCISE VIII

1. If $\mathbf{a}=5\mathbf{i}+3\mathbf{j}+4\mathbf{k}$ and $\mathbf{b}=6\mathbf{i}-8\mathbf{j}-\mathbf{k}$, find the angle between $\mathbf{a}$ nnd $\mathbf{b}$. (*D.B.S.S.C.E., 1988*)

2. If $\mathbf{a}=\mathbf{i}+2\mathbf{j}+3\mathbf{k}$, $\mathbf{b}=-2\mathbf{j}+4\mathbf{k}$, find $\mathbf{a}\cdot\mathbf{b}$. (*D.B.S.S.C.E., 1987*)

3. Find a vector of magnitude $\sqrt{51}$ which makes equal angles with the three vectors

   $\mathbf{a}=\frac{1}{3}(\mathbf{i}-2\mathbf{j}+2\mathbf{k})$, $\mathbf{b}=\frac{1}{5}(-4\mathbf{i}-3\mathbf{k})$ and $\mathbf{c}=\mathbf{k}$.
   (*Roorkee Entrance, 1987*)

4. If $\mathbf{a}=2\mathbf{i}+4\mathbf{j}-\mathbf{k}$, $\mathbf{b}=3\mathbf{i}-2\mathbf{j}+x\mathbf{k}$, then find x so that $\mathbf{a}\perp\mathbf{b}$.
   (*A.I.S.S.C.E., 1988*)

5. Find the value of λ such that the vectors $\mathbf{a}$ and $\mathbf{b}$ are perpendicular, where

   $\mathbf{a}=2\mathbf{i}+\lambda\mathbf{j}+\mathbf{k}$, and $\mathbf{b}=\mathbf{i}-2\mathbf{j}+3\mathbf{k}$ (*A.I.S.S.C.E., 1984*)

6. Two vectors $\mathbf{a}$ and $\mathbf{b}$ are such that

   $(\mathbf{a}+\mathbf{b})\cdot(\mathbf{a}-\mathbf{b})=0$. Show that $|\mathbf{a}|=|\mathbf{b}|$.
   (*A.I.S.S.C.E., 1984*)

7. If $\mathbf{a}$ and $\mathbf{b}$ are non-null vectors and $|\mathbf{a}+\mathbf{b}|=|\mathbf{a}-\mathbf{b}|$, then show that $\mathbf{a}$ and $\mathbf{b}$ are perpendicular to each other.
   (*Roorkee Entrance, 1986*)

8. If $\mathbf{a}=3\mathbf{i}+2\mathbf{j}-\mathbf{k}$, $\mathbf{b}=-2\mathbf{i}+\mathbf{j}+\mathbf{k}$, find a unit vector perpendicular to both $\mathbf{a}$ and $\mathbf{b}$. (*D.B S.S.C.E., 1989*)

9. Find $\mathbf{a}\times\mathbf{b}$ if $|\mathbf{a}|=10$, $|\mathbf{b}|=2$, $\mathbf{a}\cdot\mathbf{b}=12$.
   (*D.B.S.S.C E., 1984*)

10. If $|\mathbf{a}|=2$, $|\mathbf{b}|=5$ and $|\mathbf{a}\times\mathbf{b}|=8$, find the value of $|\mathbf{a}\cdot\mathbf{b}|$. (*D.B.S.S.C.E., 1988*)

11. Find a unit vector perpendicular to the two vectors $\mathbf{i}+\mathbf{j}$ and and $\mathbf{j}+\mathbf{k}$. (*A.I.S.S.C.E., 1987*)

12. Three vectors $\mathbf{a}=(12, 4, 3)$, $\mathbf{b}=(8, -12, -9)$, $\mathbf{c}=(33, -4, -24)$ define a parallelopiped. Evaluate the lengths of its edges, areas of its faces, and its volume.
    (*Roorkee Entrance, 1988*)

13. Find a such that the vectors $2\mathbf{i}-\mathbf{j}+\mathbf{k}$, $\mathbf{i}+2\mathbf{j}-3\mathbf{k}$, and $3\mathbf{i}+a\mathbf{j}+5\mathbf{k}$ are coplanar. (*Roorkee Entrance, 1986*)

14. If $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})=0$, and $\mathbf{b}$ and $\mathbf{c}$ are not parallel, then prove that $\mathbf{a}=\lambda\mathbf{b}+\mu\mathbf{c}$, where λ and μ are some scalars.
    (*A.I.S.S.C.E., 1989*)

15. If $\mathbf{a}=2\mathbf{i}-\mathbf{j}+\mathbf{k}$, $\mathbf{b}=\mathbf{i}+2\mathbf{j}-\mathbf{k}$, and $\mathbf{c}=2\mathbf{i}+3\mathbf{j}$, find $(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$, and $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$. (*D.B.S.S.C.E., 1989*)

16. Find the values of a for which the vectors $3\mathbf{i}+2\mathbf{j}+9\mathbf{k}$ and $\mathbf{i}+a\mathbf{j}+3\mathbf{k}$ are (*i*) perpendicular (*ii*) parallel.
    (*A.I.S.S.C.E., 1987*)

17. If $\mathbf{a}=2\mathbf{i}+5\mathbf{j}-7\mathbf{k}$, $\mathbf{b}=-3\mathbf{i}+4\mathbf{j}+\mathbf{k}$, $\mathbf{c}=\mathbf{i}-2\mathbf{j}-3\mathbf{k}$, compute $(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$ and $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ and verify that these are not the same.
18. Show that if the vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ form an orthonormal triad, then
$2\mathbf{a}=\mathbf{i}\times(\mathbf{a}\times\mathbf{i})+\mathbf{j}\times(\mathbf{a}\times\mathbf{j})+\mathbf{k}\times(\mathbf{a}\times\mathbf{k})$.
19. Prove that
$(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d})+(\mathbf{a}\times\mathbf{c})\times(\mathbf{d}\times\mathbf{b})+(\mathbf{a}\times\mathbf{d})\times(\mathbf{b}\times\mathbf{c})=2[\mathbf{b}\ \mathbf{d}\ \mathbf{c}]\mathbf{a}$.
20. Prove that
$(\mathbf{b}\times\mathbf{c})\cdot(\mathbf{a}\times\mathbf{d})+(\mathbf{c}\times\mathbf{a})\cdot(\mathbf{b}\times\mathbf{d})+(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d})=0$.

SUMMARY

1. A directed line segment is called a **vector**.
2. Two vectors are said to be **equal** if they have the same length, same or parallel supports, and the same sense,
3. A vector whose magnitude is 1 is called a **unit vector**. A vector whose initial and terminal points coincide is called the **zero vector**.
4. *Basic properties of addition of vectors and multiplication of a vector by a scalar* :

   Let **V** denote the set of all vectors and let **R** denote the set of real numbers. Then the following properties hold :

   I $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$ for all $\mathbf{a}, \mathbf{b}\in V$ (*commutativity of addition*)

   II $\mathbf{a}+(\mathbf{b}+\mathbf{c})=(\mathbf{a}+\mathbf{b})+\mathbf{c}$, for all $\mathbf{a}, \mathbf{b}, \mathbf{c}\in V$ (*Associativity of addition*)

   III $\mathbf{a}+\mathbf{0}=\mathbf{0}+\mathbf{a}$, for all $\mathbf{a}\in V$ (*Existence of zero vector*)

   IV $\mathbf{a}+(-\mathbf{a})=(-\mathbf{a})+\mathbf{a}=\mathbf{0}$, for all $\mathbf{a}\in V$. (*Existence of negative of a vector*)

   V $mn(\mathbf{a})=m(n\mathbf{a})$ for all $m, n\in R$ and $\mathbf{a}\in V$.

   VI $1(\mathbf{a})=\mathbf{a}$, for all $\mathbf{a}\in V$.

   VII $m(\mathbf{a}+\mathbf{b})=m\mathbf{a}+n\mathbf{b}$, for all $m\in R$ and $\mathbf{a}, \mathbf{b}\in V$.

   VIII $(m+n)\mathbf{a}=m\mathbf{a}+n\mathbf{a}$, for all $m, n\in R$ and $\mathbf{a}\in V$.
5. The position vector of the point dividing the join of points with position vectors **a** and **b** in the ratio $m : n$ is
$$\frac{m\mathbf{b}+n\mathbf{a}}{m+n}.$$
6. The *scalar product* of two vectors **a** and **b** inclined to each other at an angle θ is given by
$$\mathbf{a}\cdot\mathbf{b}=|\mathbf{a}|\cdot|\mathbf{b}|\cos\theta.$$
7. *Properties of scalar product* :

   If **a**, **b**, **c** be any vectors, and k be a scalar, then

   I $\mathbf{a}\cdot\mathbf{b}=\mathbf{b}\cdot\mathbf{a}$ (*Commutativity of scalar products*)

   II $\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})=\mathbf{a}\cdot\mathbf{b}+\mathbf{a}\cdot\mathbf{c}$ (*distributive property of scalar products*)

   III $(k\,\mathbf{a})\cdot\mathbf{b}=k(\mathbf{a}\cdot\mathbf{b})=\mathbf{a}\cdot(k\,\mathbf{b})$
8. The *vector product* of two vectors **a** and **b** is given by
$$\mathbf{a}\times\mathbf{b}=|\mathbf{a}|\ |\mathbf{b}|\sin\theta\ \mathbf{n},$$
where θ is the angle between the vectors **a** and **b**, **n** is a unit vector perpendicular to both **a** and **b**, the sense of **n** being such that **a**, **b**, **n** form a right-handed system.
9. *Properties of vector product* : If **a**, **b**, **c** be any three vectors, then

   (*i*) $\mathbf{a}\times\mathbf{b}=-\mathbf{b}\times\mathbf{a}$ (*anti-commutative property*)

   (*ii*) $\mathbf{a}\times(\mathbf{b}+\mathbf{c})=\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}$ (*distributive property*)

   (*iii*) $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\neq(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$, in general.

10. If **a**, **b**, **c** be a right handed system of non-coplanar vectors, then then the volume of the parallelopiped formed by **a**, **b**, **c** is given by the *scalar triple product* **a** . (**b** × **c**) and is denoted by [**a b c**].

11. A triad of vectors **a**, **b**, **c** is coplanar if and only if [**a b c**]=0.

12. For any triad of vectors **a**, **b**, **c**, **a** × (**b** . **c**)=**a** . (**b** × **c**).

13. *The vector triple product*

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=(\mathbf{a}\cdot\mathbf{c})\,\mathbf{b}-(\mathbf{a}\cdot\mathbf{b})\mathbf{c}.$$

14. If **i**, **j**, **k** be an orthonormal triad of vectors, then

$$\mathbf{i}\cdot\mathbf{i}=\mathbf{j}\cdot\mathbf{j}=\mathbf{k}\cdot\mathbf{k}=1,\ \mathbf{i}\cdot\mathbf{j}=\mathbf{j}\cdot\mathbf{k}=\mathbf{k}\cdot\mathbf{i}=0.$$

15. If

$$\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k},$$
$$\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k},$$
$$\mathbf{c}=c_1\mathbf{i}+c_2\mathbf{j}+c_3\mathbf{k},$$

then

(*i*) $|\mathbf{a}|=\sqrt{(a_1^3+a_2^2+a_3^2)}$

(*ii*) angle between **a** and $\mathbf{b}=\cos^{-1}\left(\dfrac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|}\right)$.

(*iii*) $\mathbf{a}\cdot\mathbf{b}=a_1b_1+a_2b_2+a_3b_3$

(*iv*) $\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{vmatrix}$

$$=(a_2b_3-a_3b_2)\mathbf{i}+(a_3b_1-a_1b_3)\mathbf{j}+(a_1b_2-a_2b_1)\mathbf{k}.$$

(*v*) $[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]=\begin{vmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{vmatrix}$

16. Work done=**F** . **d**, where **F** is the force vector and **d** is the displacement vector.

17. Vector moment about the origin of a force **F** acting at a point **r**=**r** × **F**.

HISTORICAL NOTE

The discovery of vectors was essentially the work of four distinguished scientists—**W.R. Hamilton** (1805-1863), **Hermann Gunther Grassman** (1809-1877), **Oliver Heaviside** (1850-1925) and **Josiah Willard Gibbs** (1839-1903).

It was in the year 1943 that the Irish mathematician Sir W.R. Hamilton discovered quaternions A year later Grassman published his *Die Aus-dehnungslehre* in which he discovered Grassman algebras. Quaternions and Grassman algebras were a sort of hyper-numbers-generalizations of complex number. Both of them were non-commutative systems. Both were however quite complicated and therefore the search for something simpler and yet equally useful continued. Around 1880 Heaviside discovered Vector Analysis to aid him in the development of electrical theory. In 1881 Gibbs published his 'Elements of Vector Analysis' and enlarged it in 1884. Vectors as we know them to-day are the same as they appeared in Heaviside's 'Electromagnetic Theory' and Gibbs' 'Elements of Vector Analysis'. Because of their simplicity, vectors have proved to be an astonishingly effective tool. They have played an ever-increasing part in physics and engineering. In fact, they have permeated all branches of science. □ □

LEONHARD EULER (1707-1783)

Leohard Euler, the most prolific mathematcian of his time, was born in Basel, in Switzerland. He was educated at the University of Basel. He was invited to the Academy at Petersberg in 1727, and stayed there upto 1741. When accepting an invitation from Frederick the Great, he came to Berlin. In 1766 he went back to Petersberg.

Euler made signal contributions in every field of mathematics which existed in his day. He wrote on the difficult topics with incredible ease, and his presentation came to be accepted as final. According to an estimate, during his lifetime, he published 530 books end papers, and at his death, he left many manuscripts, which were published by the academy, during the next 47 years. The best known of his works, the Introduction was published in 1748, and caused a revolution in analytical mathematics. The first systematic account of three dimensional geometry appeared in an appendix to this book.

CHAPTER 9

Three Dimensional Geometry

9·1. INTRODUCTION

In class XI we had studied plane geometry with the help of the analytical method. We had set up a one-to-one correspondence between ordered pairs of real numbers and points in a plane. The correspondence enabled up to look upon a geometrical figure or a region of the plane as a set of ordered pairs of real numbers. We performed algebric operations on this set and interpreted our results geometrically. We shall now apply the analytic method to the study of geometrical problem in space.

In the preceding chapter we have already seen as to how we can set up a co-ordinate system in three dimensions, and use it to define a one-to-one correspondence between points in space and ordered triples of real numbers. We have also seen as to how we can accociate with each point of the space a vector called the position vector of that point. Furthermore, we have also seen as to how we can set up a one-to-one correspondence between the set of all position-vectors with reference to a fixed origin of reference O, and the set of ordered triads of real numbers. This was done by taking a set of three mutually perpendicular lines through O as the axes of co-ordinates. We shall freely use this correspondence to translate geometrical results back and forth from vector expressions to sets of scalar expressions. It will enable us to get a feeling of the power as well as limitations of the vector method.

9·1·1. Decomposition of a Vector into Three Non-coplanar Directions i, j, k as Base Vectors.

We have already seen in Section 8·6·4 as to how we can choose a pair of three mutually perpendicular lines X′OX, Y′OY, Z′OZ, passing through a point O as the axes of co-ordinates. We have also seen in Section 8·6·3 as to how we can decompose a given vector as the sun of vectors in three non-coplanar directions. As a special case of this decomposition we saw in section 8·6·6 that by choosing a triad of unit vectors **i**, **j**, **k** along OX, OY, OZ respectively as base vectors we can decompose any vector into component vectors. The resolution turns out to be very useful in applying vectors and analytic methods together to the study of straight lines, planes, and spheres. Let us choose a fixed point O as our origin of reference for fixing up the position vectors of points in space. Also, let us choose a right-handed

system of mutually perpendicular lines X′OX, Y′OY, Z′OZ, through O as the axes of co-ordinates. Let **i**, **j**, **k** be unit vectdrs along OX, OY, OZ respectively. Also, let **r** be the position vector of P, and (x, y, z) be the cartesian co-ordinates of P. There is an obvious but important connection between **r** and (x, y, z) : If **r** is decompo‑ed ‑nto component vectors along OX, OY, and OZ, then we have

$$\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}.$$

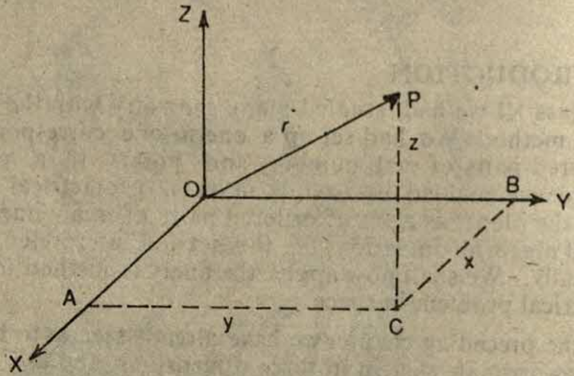


Fig. 9·1.

This relation implies that if the position vector of P(*i.e.*, **r**) is given, then we can immediately write the co-ordinates (x, y, z) of P, and conversely, if the co-ordinates (x, y, z) of P are given, then we can immediately write the position vector of P. This fact is of utmost importance because it enables us to switch over from position vectors to cartesian co-ordinates and vice-versa. Of course, in order to use this mode of transference we have to remember that the origin of reference is the same as the origin of cartesian co-ordinates. Also, the base vectors **i**, **j**, **k** for the resolution of **r** into components have been taken along the co-ordinate axes. These basic assumptions will be made throughout and we shall not repeat them everytime we use the position vector of a point and the cartesian co-ordinates of a point together. We shall often call the point P as the point **r**, as well as the point (x, y, z).

9·1·2. Direction Ratios and Directions Cosines of a Vector

Let **v** be any non-zero vector. Let its components with reference to **i**, **j**, **k**, as base vectors be a, b, c respectively, so that

$$\mathbf{v}=a\mathbf{i}+b\mathbf{j}+c\mathbf{k} \qquad ...(1)$$

If α, β, γ be the angles which **v** makes with OX, OY, OZ respectively, then

$$\mathbf{v}.\mathbf{i}=|\mathbf{v}|\cos\alpha$$
$$\mathbf{v}.\mathbf{j}=|\mathbf{v}|\cos\beta$$
$$\mathbf{v}.\mathbf{k}=|\mathbf{v}|\cos\gamma$$

Also, from (1) we find that

$$\mathbf{v}.\mathbf{i}=a,\ \mathbf{v}.\mathbf{j}=b,\ \mathbf{v}.\mathbf{k}=c, \text{ so that}$$

$$a=|\mathbf{v}|\cos\alpha,\ b=|\mathbf{v}|\cos\beta,\ c=|\mathbf{v}|\cos\gamma,$$

i.e.,
$$\frac{a}{\cos\alpha}=\frac{b}{\cos\beta}=\frac{c}{\cos\gamma}, \qquad ...(2)$$

i.e., a, b, c are proportional to $\cos\alpha, \cos\beta, \cos\gamma$ respectively. Because this reason we say that (a, b, c) are the direction ratios of the vector $\mathbf{v}$.

Also from (1), we find that

$$|\mathbf{v}|=\sqrt{(a^2+b^2+c^2)}. \qquad ...(3)$$

From (2) and (3) we find that

$$\cos\alpha=\frac{a}{\sqrt{(a^2+b^2+c^2)}}, \qquad ...(4)$$

$$\cos\beta=\frac{b}{\sqrt{(a^2+b^2+c^2)}},$$

$$\cos\gamma=\frac{c}{\sqrt{(a^2+b^2+c^2)}}.$$

The numbers $\cos\alpha, \cos\beta, \cos\gamma$ are called the direction cosines of $\mathbf{v}$. *The direction cosines of a vector are the cosines of the angles which the vector makes with the co-ordinate axes OX, OY, and OZ respectively.* It is usual to write the direction cosines of a vector as an ordered triad $(\cos\alpha, \cos\beta, \cos\gamma)$. The letters l, m, n are generally used to denote $\cos\alpha, \cos\beta, \cos\gamma$ respectively. We shall often come across the statement (l, m, n) are the direction cosines of the vector $\mathbf{v}$. This statement will mean that if $\mathbf{v}$ makes angles α, β, γ with OX, OY, OZ respectively, then

$$\cos\alpha=l,\ \cos\beta=m,\ \cos\gamma=n.$$

From (4), we find that

$$\cos^2\alpha+\cos^2\beta+\cos^2\gamma=1.$$

In other words, if (l, m, n) be the direction cosines of a vector, then

$$\boxed{l^2+m^2+n^2=1} \qquad ...(A)$$

Also, from (4) we find that

$$\boxed{\begin{aligned} l&=a/\sqrt{(a^2+b^2+c^2} \\ m&=b/\sqrt{(a^2+b^2+c^2)} \\ n&=c/\sqrt{(a^2+b^2+c^2)} \end{aligned}} \qquad ...(B)$$

9·13. Angle between Two Vectors whose Direction Cosines are given :

Let **v** and **u** be two vectors whose direction cosines with respect to a set of co-ordinate axes X′OX, Y′OY, Z′OZ are

$$(l_1,\ m_1,\ n_1)$$

and $(l_2,\ m_2,\ n_2)$ respectively.

$$\mathbf{u}.\mathbf{i}=|\mathbf{u}|\ l_1,\ \mathbf{u}.\mathbf{j}=|\mathbf{u}|\ m_1,\ \mathbf{u}.\mathbf{k}=|\mathbf{u}|\ n_1,$$

$$\frac{1}{|\mathbf{u}|}\mathbf{u}=l_2\mathbf{i}+m_2\mathrm{j}+n_2\mathbf{k}. \qquad ...(1)$$

Similarly,

$$\frac{1}{|\mathbf{v}|}\mathbf{v}=l_3\mathbf{i}+m_3\mathbf{j}+n_3\mathbf{k}. \qquad ...(2)$$

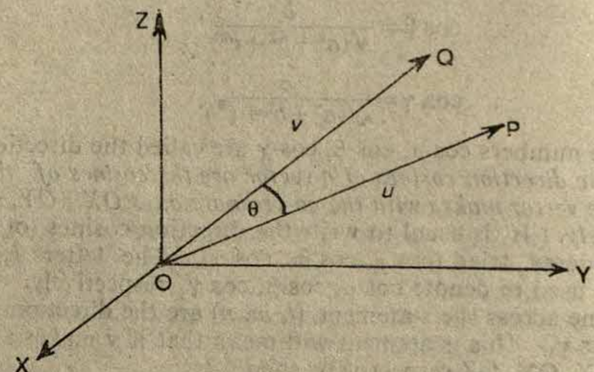


Fig. 9·2.

From (1) and (2) we find that if θ be the angle between the vectors **u** and **v**, then

$$\cos\theta=\frac{\mathbf{u}\ .\ \mathbf{v}}{|\mathbf{u}|\ |\mathbf{v}|}$$

$$=\frac{\mathbf{u}}{|\mathbf{u}|}\cdot\frac{\mathbf{v}}{|\mathbf{v}|}$$

$$=(l_1\mathbf{i}+m_1\mathrm{j}+n_1k).(l_2\mathbf{i}+m_2\mathbf{j}+n_2\mathbf{k})$$

$$=l_1l_2+m_1m_2+n_1n_2.$$

Thus, we find that if θ be the angle between two vectors with direction cosines $(l_1,\ m_1,\ n_1)$ and $(l_2,\ m_2,\ n_2)$ then

$$\boxed{\cos\theta=l_1l_2+m_1m_2+n_1n_2} \qquad ...(A)$$

If (a_1, b_1, c_1), (a_2, b_2, c_2) be the direction ratios of two vectors **u** and **v** respectively, and (l_1, m_1, n_1), (l_2, m_2, n_2) be their direction cosines, then

$$l_1 = a_1/\sqrt{(a^2{}_1+b^2{}_1+c^2{}_1)},$$
$$m_1 = b_1/\sqrt{(a^2{}_1+b^2{}_1+c^2{}_1)},$$
$$n_1 = c_1/\sqrt{(a^2{}_1+b^2{}_1+c^2{}_1)},$$
$$l_2 = a_2/\sqrt{(a_2{}^2+b_2{}^2+c^2{}_2)},$$
$$m_2 = b_2/\sqrt{(a^2{}_2+b_2{}^2+c_2{}^2)},$$
$$n_2 = c_2/\sqrt{(a_2{}^2+b_2{}^2+c^2{}_2)},$$

so that from (A) above we have

$$\cos\theta = \frac{a_1a_2+b_1b_2+c_1c_2}{\sqrt{((a^2{}_1+b^2{}_1+c_1{}^2)}\ \sqrt{(a_2{}^2+b_2{}^2+c^2{}_2)}}.$$

Example 1. *Find the angle between the vectors with direction ratios (3, 4, 5) and (4, −3, 5).*

Solution. The angle θ between the vectors is given by

$$\cos\theta = \frac{3.4+4.(-3)+5.5}{\sqrt{(3^2+4^2+5^2)}\ \sqrt{(4^2+(-3)^2+5^2)}}$$
$$= \frac{25.25}{(5\sqrt{2})(5\sqrt{2})} = \frac{1}{2}.$$

Therefore the vectors are inclined to each other at the angle $\pi/3$.

EXERCISE 9 (*a*)

1. Find the direction cosines of the vector with direction ratios :
   (*a*) (2, 1, 2) (*b*) (−2, 1, 2)
   (*c*) (2, −1, 2) (*d*) (2, −2, 1).
2. Find the direction cosines of the vector with direction ratios :
   (*a*) (−3, 4, 5) (*b*) (5, −4, 3)
   (*c*) (−5, 3, 4) (*d*) (3, −5, 4).
3. What are the direction cosines of the vectoss :
   (*a*) **i** (*b*) **j**
   (*c*) **k** (*d*) −**i**
   (*f*) −**j** (*g*) −**k** ?
4. Find the angle between the vectors whose direction cosines are :
   (*a*) (1, 0, 0) and $(1/\sqrt{2}, 1/\sqrt{2}, 0)$,
   (*b*) $(1/\sqrt{3}, 1/\sqrt{3}, -1/\sqrt{3})$, and $(1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3})$,

9·2. DISTANCE BETWEEN TWO POINTS

Let P (x_1, y_1, z_1), Q (x_2, y_2, z_2) be two given points. The position vectors of P and Q with respect to 0 as the origin of reference are

$$\overrightarrow{OP}=x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k},$$

$$\overrightarrow{OQ}=x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k},$$

respectively.

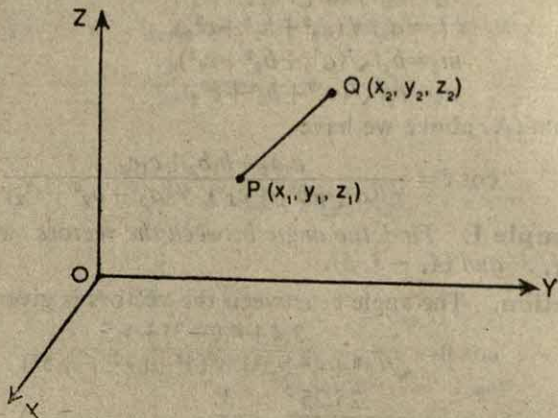


Fig. 9·3.

$$\begin{aligned}PQ&=|\overrightarrow{PQ}|\\&=|\overrightarrow{OQ}-\overrightarrow{OP}|\\&=|(x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k})-(x_1\mathbf{i}+y_1\mathbf{j}+z_2\mathbf{k})|\\&=|(x_2-x_1)\,\mathbf{i}+(y_2-y_1)\,\mathbf{j}+(z_2-z_1)\,\mathbf{k}|\\&=\sqrt{\{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2\}}.\end{aligned}$$

Thus we find that *the distance d between the points* (x_1, y_1, z_1) and (x_2, y_2, z_2) *is given by*

$$\boxed{d=\sqrt{\{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2\}}}$$

Corollary. The distance d of the point (x, y, z) from the origin is

$$d=\sqrt{x^2+y^2+z^2}$$

Proof. $d=\sqrt{\{(x-0)^2+(y-0)^2+(z-0)^2\}},$

$$=\sqrt{(x^2+y^2+z^2)}.$$

Example 2. *Show that the points $(9, 1, -3)$, $(3, 1, 3)$ and $(1, -1, -5)$ are the vertices of an equilateral triangle.*

Solution. Let the co-ordinates of the vertices A, B, C of the triangle ABC be (9, 1, −3), (3, 1, 3) and (1, −1, −5) respectively.

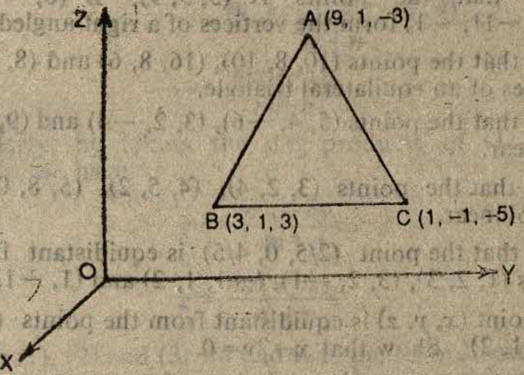


Fig. 9·4.

$$BC=\sqrt{\{(1-3)^2+(-1-1)^2+(-5-3)^2\}}=\sqrt{72},$$
$$CA=\sqrt{\{(1-9)^2+(-1-1)^2+(-5-(-3))^2\}}=\sqrt{72},$$
$$AB=\sqrt{\{(3-9)^2+(1-1)^2+(3-(-3))^2\}}=\sqrt{72}.$$

Since BC=CA=AB, therefore the ΔABC is equilateral.

Example 3. *Show that the points (4, 5, 6), (2, 2, 2) and (6, 8, 10) are collinear.*

Solution. Let us name the points (4, 5, 6), (2, 2, 2) and (6, 8, 10) as A, B, and C respectively.

$$AB=\sqrt{\{(2-4)^2+(2-5)^2+(2-6)^2\}}=\sqrt{29},$$
$$BC=\sqrt{\{(6-2)^2+(8-2)^2+(10-2)^2\}}=\sqrt{116}=2\sqrt{29},$$
$$AC=\sqrt{\{(6-4)^2+(8-5)^2+(10-6)^2\}}=\sqrt{29}.$$

Since AB+AC=BC, therefore the points A, B and C are collinear.

EXERCISE 9 (*b*)

1. Find the distance of each of the following points from the origin :
   (*a*) (3, 4, 12), (9, −12, −8), (−3, 2, 5)
   (*b*) (−4, −2, 10), (3, −5, 4), (8, 1, −2)
2. Find the distance between each of the following pairs of points :
   (*a*) (3, −7, 8), (−4, 6, 2) ;
   (*b*) (5, 1, 6), (−2, −3, 4) ;
   (*c*) (1, −3, 4), (4, −2, −3).

3. Show that the points A (4, 6, −5), B (0, 2, 3), and C (−4, −4, −1) from the vertices of an isosceles triangle.

4. Show that the points A (5, 9, 9), B (0, −1, −6), and C (5, −11, −1) form the vertices of a right angled triangle.

5. Show that the points (10, 8, 10), (16, 8, 6) and (8, 6, 2) are the vertices of an equilateral triangle.

6. Show that the points (5, 4, −6), (3, 2, −4) and (9, 8, −10) are collinear.

7. Show that the points (3, 2, 4), (4, 5, 2), (5, 8, 0), (2, −1, 6) are collinear.

8. Show that the point (2/5, 0, 4/5) is equidistant from the four points (1, 2, 3), (3, 2, −1), (−1, 1, 2) and (1, −1, −2).

9. The point (x, y, z) is equidistant from the points (1, 3, 2) and (3, −1, 2). Show that $x-2y=0$.

9·3. SECTION FORMULA

Let P (x_1, y_1, z_1) and Q (x_2, y_2, z_2) be two given points and let R (x, y, z) be the point which divides the join of PQ in the ratio $n : m$.

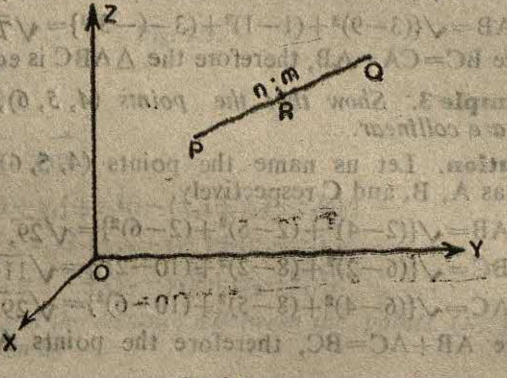


Fig. 9·5.

With respect to O as the origin of reference, the position vectors of P, Q and R are

$$\mathbf{r}_1=x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k},$$
$$\mathbf{r}_2=x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k},$$
$$\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}.$$

From our study of vectors we know that since R divides PQ in the ratio $n : m$, therefore

$$\mathbf{r}=\frac{m\mathbf{r}_1+n\mathbf{r}_2}{m+n}, \qquad \ldots(1)$$

Taking dot products of both sides with **i**, we have

$$\mathbf{r} \cdot \mathbf{i} = \frac{m\mathbf{r}_1 + n\mathbf{r}_2}{m+n} \cdot \mathbf{i}$$

$$= \frac{m(\mathbf{r}_1 \cdot \mathbf{i}) + n(\mathbf{r}_2 \cdot \mathbf{i})}{m+n},$$

or
$$x = \frac{mx_1 + nx_2}{m+n}. \qquad \ldots(2)$$

Similarly, by taking the dot products of (1) with **j** and **k** successively, we have

$$y = \frac{my_1 + ny_2}{m+n},$$

$$z = \frac{mz_1 + nz_2}{m+n}. \qquad \ldots(3)$$

From (1), (2) and (3) we find that the co-ordinates of the point which divides the join of (x_1, y_1, z_1) and (x_2, y_2, z_2) in the ratio $n : m$ are given by

$$\boxed{\begin{aligned} x &= \frac{mx_1 + nx_2}{m+n}, \\ y &= \frac{my_1 + ny_2}{m+n}, \\ z &= \frac{mz_1 + nz_2}{m+n}. \end{aligned}}$$

Remarks 1. The above discussion and result holds whether R divides PQ internally or externally. In the case of internal division, both m and n are taken to be positive. In the case of external division, m and n are taken to be of opposite signs. It is usual to taken n as positive and m as negative (though it really does not matter even if we take n as negative and m as positive).

2. $m+n$ cannot be zero, that is, we cannot have $m : n = 1 : -1$. Geometrically speaking it only means that we cannot divide a line segment *externally* in the ratio $1 : 1$.

3. Putting $m = n = 1$, we find that the co-ordinates of the mid-point of the join of (x_1, y_1, z_1) and (x_2, y_2, z_2) are $(\frac{1}{2}(x_1+x_2), \frac{1}{2}(y_1+y_2), \frac{1}{2}(z_1+z_2))$.

Example 4. *Find the ratio in which the yz-plane divides the line joining the points $(-2, 4, 7)$ and $(3, -5, 8)$.*

Solution. Suppose that the yz-plane divides the join of $(-2, 4, 7)$ and $(3, -5, 8)$ in the ratio $\lambda : 1$. The co-ordinates (x, y, z) of this point are given by

$$x=(-2\cdot 1+3\cdot\lambda)/(\lambda+1),$$
$$y=(4.1+(-5)\,.\lambda)/(\lambda+1),$$
$$z=(7.1+8.\lambda)/(\lambda+1).$$

If the point (x, y, z) lies on the yz-plane, then its x-co-ordinate must be zero.

$\therefore \quad (-2+3\lambda)/(\lambda+1)=0,$

or $\quad -2+3\lambda=0,$

or $\quad \lambda=\frac{2}{3}.$

The desired ratio $=\frac{2}{3}:1=2:3.$

Example 5. *Show that the points $A\ (3, 2, -4)$, $B\ (5, 4, -6)$ and $C\ (9, 8, -10)$ are collinear.*

Solution. The co-ordinates of any point on the join of A (3, 2, −4) and B (5, 4, −6) are

$$\left(\frac{5\lambda+3}{\lambda+1}, \frac{4\lambda+2}{\lambda+1}, \frac{-6\lambda-4}{\lambda+1}\right)$$

If A, B, C are collinear, then there must be some value of λ for which

$$\left(\frac{5\lambda+3}{\lambda+1}, \frac{4\lambda+2}{\lambda+1}, \frac{-6\lambda-4}{\lambda+1}\right)=(9, 8, -10),$$

or $$\frac{5\lambda+3}{\lambda+1}=9,\ \frac{4\lambda+2}{\lambda+1}=8,\ \frac{-6\lambda-4}{\lambda+1}=-10. \qquad ...(1)$$

If A, B, C are collinear, *all* the relations (*i*) must be satisfied by the same value of λ. The first of these relations gives

$$5\lambda+3=9\ (\lambda+1)$$

or $$\lambda=-\tfrac{3}{2}.$$

By actual substitution we find that $\lambda=-\dfrac{3}{2}$ satisfies the remaining two relations also. Hence the points A, B, C are collinear.

Example 6. *Given that the points*

$$P(-2, 3, 5),\ Q\ (1, 2, 3),\ R\ (7, 0, -1)$$

are collinear, find the ratio in which R divides the join of P and Q.

Solution. Let R divide the join of P and Q in the ratio $\lambda : 1$. The co-ordinates of the point dividing the join of P and Q in the ratio $\lambda : 1$ are

$$\left(\frac{\lambda-2}{\lambda+1}, \frac{2\lambda+3}{\lambda+1}, \frac{3\lambda+5}{\lambda+1}\right).$$

If this is the same as the point (7, 0, 1), we have

$$\frac{\lambda-2}{\lambda+1}=7,\ \frac{2\lambda+3}{\lambda+1}=0,\ \frac{3\lambda+5}{\lambda+1}=-1 \qquad ...(i)$$

The first of the relations (*i*) gives $\lambda=-\frac{3}{2}$. Therefore the required ratio $=-\frac{3}{2}:1=3:-2$. Hence $(7, 0, -1)$ divides the join of $(-2, 3, 5)$, and $(1, 2, 3)$ externally in the ratio $3:2$.

Note : The second (or the third) of the relations (*i*) would also have given the same value of λ. In fact, the correctness of the result can always be seen by checking that all the three relations (*i*) give the same value of λ.

Example 7. *Show that the lines joining the vertices of a tetrahedron to the centroids of the opposite faces meet in a point.*

Solution. Let A (x_1, y_1, z_1), B (x_2, y_2, z_2), C (x_3, y_3, z_3), D (x_4, y_4, z_4) be the vertices of a tetrahedron. The coordinates of the mid-point L of BC are

$$\left(\frac{x_2+x_3}{2}, \frac{y_2+y_3}{2}, \frac{z_2+z_3}{2}\right).$$

The centroid G_1 of the Δ ABC divides AL in the ratio $2:1$. Therefore the co-ordinates of G_1 are

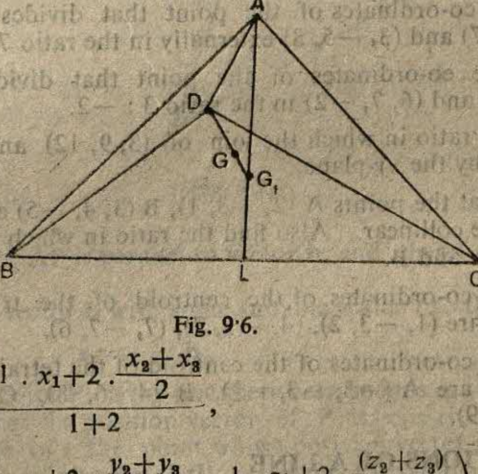


Fig. 9·6.

$$\left(\frac{1 \cdot x_1+2 \cdot \frac{x_2+x_3}{2}}{1+2}, \frac{1 \cdot y_1+2 \cdot \frac{y_2+y_3}{2}}{1+2}, \frac{1 \cdot z_1+2 \cdot \frac{(z_2+z_3)}{2}}{1+2}\right)$$

i.e., $\left(\frac{1}{3}(x_1+x_2+x_3), \frac{1}{3}(y_1+y_2+y_3), \frac{1}{3}(z_1+z_2+z_3)\right)$.

The co-ordinates of the point G which divides DG_1 in the ratio $3:1$ are

$$\left(\frac{3 \cdot \frac{1}{3}(x_1+x_2+x_3)+x_4}{3+1}, \frac{3 \cdot \frac{1}{3}(y_1+y_2+y_3)+y_4}{3+1}, \frac{3 \cdot \frac{1}{3}(z_1+z_2+z_3)+z_4}{3+1}\right)$$

i.e., $\left(\dfrac{x_1+x_2+x_3+x_4}{4}, \dfrac{y_1+y_2+y_3+y_4}{4}, \dfrac{z_1+z_2+z_3+z_4}{4}\right)$

Since the expressions $\frac{1}{4}(x_1+x_2+x_3+x_4)$, $\frac{1}{4}(y_1+y_2+y_3+y_4)$, $\frac{1}{4}(z_1+z_2+z_3+z_4)$ remain the same when the suffixes 1, 2, 3, 4 are cyclically interchanged, therefore the point

$$G\left(\frac{x_1+x_2+x_3+x_4}{4}, \frac{y_1+y_2+y_3+y_4}{4}, \frac{z_1+z_2+z_3+z_4}{4}\right)$$

also lies on the join of A (resp. B, C) to the centroid of ΔBCD (resp. ΔCDA, ΔDAB), and divides it in the ratio 3 : 1.

Hence the lines joining the vertices of the tetrahedron ABCD to the centroids of the opposite faces are concurrent.

EXERCISE 9 (*c*)

1. Find the mid-point of the join of A (3, −4, 2) and B (−7, 6, −8).
2. Find the co-ordinates of the point that divides the join (2, −3, 1) and (3, 4, −5) internally in the ratio 1 : 3.
3. Find the co-ordinates of the point that divides the join of (−2, 4, 7) and (3, −5, 8) externally in the ratio 7 : 8.
4. Find the co-ordinates of the point that divides the join of (5, 0, 4) and (6, 7, −2) in the ratio 3 : −2.
5. Find the ratio in which the join of (3, 9, 12) and (8, 0, 9) is divided by the xy-plane.
6. Show that the points A (2, −3, 1), B (3, 4, −5) and C (5, 18, −17) are collinear. Also find the ratio in which C divides the join of A and B.
7. Find the co-ordinates of the centroid of the triangle whose vertices are (1, −3, 2), (4, −5, 7), (7, −7, 6).
8. Find the co-ordinates of the centroid of the tetrahedron whose vertices are A (−5, −3, −2), B (4, 6, 8), C (2, 1, −7), D (3, 4, 9).

9·4. EQUATIONS OF A LINE

The position vectors of points on a given line can be expressed in terms of some fixed vectors and a variable scalar, called parameter, such that for arbitrary values of the parameter the resulting position vector represents points on the line, and conversely, the position vector of each point on the locus arises for some suitable values of the parameters. Such a relation is called a *parametric vectorial equation* of the line. We shall obtain the parametric vectorial equation of a line in two different forms and also find the equivalent cartesian equations.

9·4·1. Parametric Vectorial Equation of a line

To find the vectorial equation of a line which passes through a given point and is parallel to a given line.

Take any point O as the origin of reference. Let $\mathbf{r}_1$ be the position vector of the given point and let $\mathbf{b}$ be *any* vector parallel to the given line. Let $\mathbf{r}$ be the position vector of any point P on the given line.

We have

$$\mathbf{r}=\overrightarrow{OP},$$

$$=\overrightarrow{OA}+\overrightarrow{AP},$$

$$=\mathbf{r}_1+\overrightarrow{AP}.$$

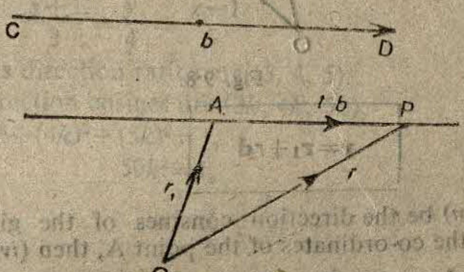


Fig. 9·7.

The vector $\overrightarrow{AP}$ is parallel to vector $\mathbf{b}$, and therefore it is the product of the vector $\mathbf{b}$ with some suitable scalar t so that $\overrightarrow{AP}=t\mathbf{b}$.

$$\therefore \quad \mathbf{r}=\mathbf{r}_1+t\,\mathbf{b}.$$

To each point P on the line there corresponds a value of the scalar t such that the position vector of P is $\mathbf{r}_1+t\mathbf{b}$. Conversely, for each value of t the point whose position vector is $\mathbf{r}_1+t\mathbf{b}$ lies on the line. Therefore $\mathbf{r}=\mathbf{r}_1+t\mathbf{b}$ represents the parametric vectorial equation of the line through $\mathbf{r}_1$ parallel to the vector $\mathbf{b}$.

In the above discussion suppose that the unit vector along $\mathbf{b}$ is $\overrightarrow{CE}=\mathbf{d}$ say, so that $\mathbf{b}=|\mathbf{b}|\,\mathbf{d}$. Then $\text{AP}=r\,\mathbf{d}$, where $\overrightarrow{AP}=r$ [r is positive for points on the ray parallel to $\mathbf{d}$ and starting from A, and negative for points on the ray parallel to $\mathbf{d}$ and having the sense opposite to $\mathbf{d}$ (shown as dotted line in Fig. 9·8)].

Then

$$\mathbf{r}=\overrightarrow{OP}=\overrightarrow{OA}+\overrightarrow{AP}$$
$$=\mathbf{r}_1+r\mathbf{d},$$

and as above, the parametric vectorial equation of the line is given by

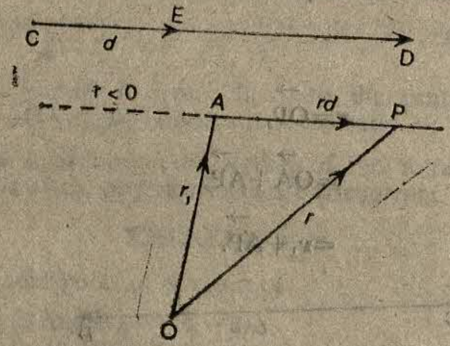


Fig. 9·8.

$$\boxed{\mathbf{r}=\mathbf{r}_1+r\mathbf{d}} \quad ...(iv)$$

If (l, m, n) be the direction consines of the given line, and (x_1, y_1, z_1) be the co-ordinates of the point A, then (iv) becomes

$$x\mathbf{i}+y\mathbf{j}+z\mathbf{k}=x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k}+r\,(l\mathbf{i}+m\mathbf{j}+n\,\mathbf{k}) \quad ...(v)$$

By taking dot products of both sides of with $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ in succession, we have

$$\boxed{x=x_1+lr,\ y=y_1+mr,\ z=z_1+nr} \quad ...(vi)$$

Relations (vi) give the cartesian co-ordinates of any point P on the line in terms of its distance r from A, distanee r is being measured here algebraically. It is positive for points on one side of A and negative for points on the other side. From (vi) we find that *the equations of the line through* (x_1, y_1, z_1) *and having direction cosines* (l, m, n) *are given* by

$$\boxed{\frac{x-x_1}{l}=\frac{y-y_1}{m}=\frac{z-z_1}{n}=r} \quad ...(vii)$$

Remark. The above equations are sometimes referred to as *equations of a line in point-distance form.*

Example 8. *Find the equations of the line passing through the point (1, 2, 3) and parallel to the line*

$$\frac{x-6}{12}=\frac{y-2}{4}=\frac{z+7}{5}.$$

Solution. The desired line has direction ratios (12, 4, 5). Also it passes through the point (1, 2, 3). Therefore its equation is

$$\frac{x-1}{12}=\frac{y-2}{4}=\frac{z-3}{5}.$$

Example 9. *Find the co-ordinates of the points distant $10\sqrt{2}$ units from the point (2, 3, 4) and lying on the line*

$$\frac{x+1}{3}=\frac{y}{4}=\frac{z-1}{5}.$$

Solution. Since the desired line is parallel to

$$\frac{x+1}{3}=\frac{y}{4}=\frac{z-1}{5}.$$

therefore its direction ratios are (3, 4, 5).

Its direction cosines are $(3k, 4k, 5k)$,

where $(3k)^2+(4k)^2+(5k)^2=1$,

or $$50k^2=1,$$

so that $$k=\pm\frac{1}{5\sqrt{2}}.$$

$\therefore$ The direction cosines are $\pm\dfrac{1}{5\sqrt{2}}(3, 4, 5)$.

The equations of the line through the point (2, 3, 4) and having direction cosines $\dfrac{1}{5\sqrt{2}}(3, 4, 5)$ are

$$\frac{x-2}{3/5\sqrt{2}}=\frac{y-3}{4/5\sqrt{2}}=\frac{z-4}{5/5\sqrt{2}}=r, \qquad ...(i)$$

The co-ordinates of the two points on the line (*i*) at a distance $10\sqrt{2}$ units are obtained by putting $r=\pm 10\sqrt{2}$ in (*i*). Therefore the desired co-ordinates are given by

$$\left(2+\frac{3}{5\sqrt{2}}r,\ 3+\frac{4}{5\sqrt{2}}r,\ 4+\frac{5}{5\sqrt{2}}r\right),$$

where $$r=\pm 10\sqrt{2},$$

i.e., $(2\pm 6, 3\pm 8, 4\pm 10)$.

Therefore the points are (8, 11, 14) and (−4, −5, −6).

9·4 2. Equations of a Line Through Two Given Points

Let $\mathbf{r}_1$, $\mathbf{r}_2$ be the position vectors of the two given points A_1, A_2 with reference to a point O as origin of reference. Then the line A_1A_2 is parallel to the vector

$$\overrightarrow{A_1A_2}=\overrightarrow{OA_2}-\overrightarrow{OA_1}$$
$$=\mathbf{r}_2-\mathbf{r}_1.$$

The vector equation of the line A_1A_2

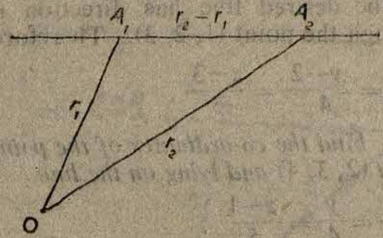


Fig. 9·9.

is

$$\mathbf{r}=\mathbf{r}_1+t\,(\mathbf{r}_2-\mathbf{r}_1)=(1-t)\,\mathbf{r}_2+t\,\mathbf{r}_1.$$

Thus we find that *the parametric vectorial equation of the lin through the points having pasition vectors* $\mathbf{r}_1$ and $\mathbf{r}_2$ is

$$\boxed{\mathbf{r}=\mathbf{r}_1+t\,(\mathbf{r}_2-\mathbf{r}_1)}$$

Corollary. If the cartesian co-ordinates of A_1 and A_2 be (x_1, y_1, z_1) and (x_2, y_2, z_2),

then $\quad \mathbf{r}_1=x_1\,\mathbf{i}+y_1\,\mathbf{j}+z_1\,\mathbf{k},\ r_2=x_2\,\mathbf{i}+y_2\,\mathbf{j}+z_2\,\mathbf{k}.$

Therefore the equation of the line $A_1\,A_2$ is given by

$$\mathbf{r}-\mathbf{r}_1=t\,(\mathbf{r}_2-\mathbf{r}_1),$$

or $\quad (x-x_1)\,\mathbf{i}+(y-y_1)\,\mathbf{j}+(z-z_1)\,\mathbf{k}=t\,[(x_2-x_1)\,\mathbf{i}+(y_2-y_1)\,\mathbf{j}+(z_2-z_1)\,\mathbf{k}],$

so that

$$x-x_1=t\,(x_2-x_1),\ y-y_1=t\,(y_2-y_1),\ z-z_1=t\,(z-z_2),$$

or

$$\boxed{\frac{x-x_1}{x_2-x_1}=\frac{y-y_1}{y_2-y_1}=\frac{z-z_1}{z_2-z_1}=t}$$

Example 10. *Find the equations of the line passing through the points* (1, ?, −1) *and* (3, −1, 2). *At what point does it meet the yz-plane ?*

Solution. Take $(x_1, y_1, z_1)=(1, 2, -1)$ and $(x_2, y_2, z_2)=(3, -1, 2)$. The desired equations are

$$\frac{x-1}{3-1}=\frac{y-2}{-1-2}=\frac{z-(-1)}{2-(-1)},$$

or
$$\frac{x-1}{2}=\frac{y-2}{-3}=\frac{z+1}{3}. \qquad ...(i)$$

Since the x-co-ordinate of every point lying on the yz-plane is 0, therefore y and z co-ordinates of such a point are obtained by putting $x=0$ in (i).

$$\therefore \quad \frac{0-1}{2}=\frac{y-2}{-3}=\frac{z+1}{3},$$

or
$$y=\frac{7}{2},\ z=\frac{-5}{2}.$$

$\therefore$ The line meets the yz-plane in the point $\left(0, \frac{7}{2}, \frac{-5}{2}\right)$.

EXERCISE 9 (*d*)

1. Find the vector equation of the line passing through the point whose position vector is $2\mathbf{i}+3\mathbf{j}-\mathbf{k}$, and parallel to the vector $\mathbf{i}-\mathbf{j}+2\mathbf{k}$.
2. Find the vector equation of the line passing through the point $(1, -1, 2)$ and having direction ratios $(2, 1, 2)$.
3. Find the cartesian equation of the line passing through the point $(-1, 1, 1)$ and having direction ratios $(1, 2, -2)$.
4. Find the vector equation of the line whose cartesian equations are
$$\frac{x-1}{2}=\frac{y+1}{-3}=\frac{z+2}{4}.$$
5. Find the cartesian equations of the line whose vector equation is
$$\mathbf{r}=2\mathbf{i}+3\mathbf{j}+t(\mathbf{j}-2\mathbf{k}).$$
6. Find the vector equation of the line passing through the points $\mathbf{i}-\mathbf{j}+\mathbf{k}$ and $2\mathbf{i}+\mathbf{j}-3\mathbf{k}$.
7. Find the vector equations of the line passing through the points $(2, 1, 3)$ and $(-1, 2, -5)$.
8. Find the cartesian equations of the line passing through the points $(2, -1, -3)$ and $(-3, 1, 4)$.
9. Find the equations of the line joining the points $(-1, -1, 2)$ and $(3, 1, 3)$.
10. Find the cartesian equation of the line whose vector equation is
$$\mathbf{r}=(\mathbf{i}-\mathbf{j}+\mathbf{k})(1-t)+(2\mathbf{i}-3\mathbf{j}+4\mathbf{k})t.$$

9·4·3. Angle Between Two Lines

Let $$\mathbf{r}=\mathbf{r}_1+t\mathbf{b}_1, \qquad \text{...(1)}$$

and $$\mathbf{r}=\mathbf{r}_2+s\mathbf{b}_2, \qquad \text{...(2)}$$

be the equations of two lines, where t and s are parameters. These straight lines are in the directions determined by the vectors $\mathbf{b}_1$ and $\mathbf{b}_2$. The angle θ between these lines is defined to be the angle between the directions $\mathbf{b}_1$ and $\mathbf{b}_2$.

Since $\mathbf{b}_1 \cdot \mathbf{b}_2=|\mathbf{b}_1| \cdot |\mathbf{b}_2| \cos\theta$,
therefore θ is given by

$$\cos\theta=\frac{\mathbf{b}_1 \cdot \mathbf{b}_2}{|\mathbf{b}_1|\ |\mathbf{b}_2|}. \qquad \text{...(3)}$$

If the equations of the two lines are given in cartesian form as

$$\frac{x-x_1}{a_1}=\frac{y-y_1}{b_1}=\frac{z-z_1}{c_1}, \qquad \text{...(4)}$$

and $$\frac{x-x_2}{a_2}=\frac{y-y_2}{b_2}=\frac{z-z_2}{c_2}, \qquad \text{...(5)}$$

so that

$$\mathbf{b}_1=a_1\mathbf{i}+b_1\mathbf{j}+c_1\mathbf{k},$$
$$\mathbf{b}_2=a_2\mathbf{i}+b_2\mathbf{j}+c_2\mathbf{k},$$

then

$$\mathbf{b}_1 \cdot \mathbf{b}_2=a_1a_2+b_1b_2+c_1c_2,$$
$$\mathbf{b}_1=\sqrt{(a_1^2+b_1^2+c_1^2)},$$
$$\mathbf{b}_2=\sqrt{(a_2^2+b_2^2+c_2^2)},$$

so that

$$\cos\theta=\frac{a_1a_2+b_1b_2+c_1c_2}{\sqrt{(a_1^2+b_1^2+c_1^2)}\sqrt{(a_2^2+b_2^2+c_2^2)}}. \qquad \text{...(6)}$$

If the equations of the lines are

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1}, \qquad \text{...(7)}$$

$$\frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}=\frac{z-z_2}{n_2}, \qquad \text{...(8)}$$

where (l_1, m_1, n_1) and (l_2, m_2, n_2) are direction cosines instead of direction ratios, then

$$l_1^2+m_1^2+n_1^2=1,\ l_2^2+m_2^2+n_2^2=1,$$

and the formula (6) gives

$$\boxed{\cos\theta=l_1l_2+m_1m_2+n_1n_2}$$

Corollary. The lines (4) and (5) are perpendicular to each other if and only if

$$a_1a_2+b_1b_2+c_1c_2=0. \quad ...(9)$$

Similarly, the lines (7) and (8) are perpendicular to each other if and only if

$$l_1l_2+m_1m_2+n_1n_2=0. \quad ...(10)$$

Remark. The lines (4) and (5) are parallel if and only if

$$\frac{a_1}{a_2}=\frac{b_1}{b_2}=\frac{c_1}{c_2}.$$

Equivalently, the lines (7) and (8) are parallel if and only if

$$\frac{l_1}{l_2}=\frac{m_1}{m_2}=\frac{n_1}{n_2}.$$

Example 11. *Find the angle between the lines*

$$\mathbf{r}=\mathbf{i}+s(2\mathbf{i}-2\mathbf{j}+\mathbf{k}),$$

and

$$\mathbf{r}=2\mathbf{j}+t(\mathbf{i}+2\mathbf{j}-2\mathbf{k}).$$

Solution. The angle θ between the given lines is the same as the angle between the vectors

$$\mathbf{a}=2\mathbf{i}-2\mathbf{j}+\mathbf{k},\ \mathbf{b}=\mathbf{i}+2\mathbf{j}-2\mathbf{k}.$$

$\therefore \quad \cos\theta=(\mathbf{a}\,.\,\mathbf{b})/\,|\,\mathbf{a}\,|\,.\,|\,\mathbf{b}\,| \quad ...(1)$

Now $\quad \mathbf{a}\,.\,\mathbf{b}=2\,.\,1+(-2)\,.\,2+1\,.\,(-2)=-4.$

$$|\,\mathbf{a}\,|=\sqrt{2^2+(-2)^2+1^2}=3,$$

$$|\,\mathbf{b}\,|=\sqrt{1^2+2^2+(-2)^2}=3.$$

$\therefore$ From (1), we have

$$\cos\theta=\frac{-4}{3.3}=\frac{-4}{9}.$$

Hence the required angle is $\cos^{-1}\left(\frac{-4}{9}\right).$

Remark. The acute angle between the lines is $\cos^{-1}\left(\frac{4}{9}\right).$

Example 12. *Find the equations of the perpendicular drawn from the point (2, 4, −1) to the line*

$$\frac{x+5}{1}=\frac{y+3}{4}=\frac{z-6}{-9}$$

and obtain the co-ordinates of the foot of the perpendicular.

(*D.B.S.S.C.E.*, *1988*)

Solution. The co-ordinates of any point Q on

$$\frac{x+5}{1}=\frac{y+3}{4}=\frac{z-6}{-9}=r \quad ...(1)$$

are $(-5+r, -3+4r, 6-9r)$. Q is the foot of the perpendicular drawn from the point P(2, 4, −1) provided PQ is perpendicular to (1). Now the direction ratios of PQ are $(-5+r-2, -3+4r-4, 6-9r+1)$,

i.e., $$(-7+r, -7+4r, 7-9r). \qquad ...(2)$$

PQ is perpendicular to the line (1) provided

$$(-7+r)+4(-7+4r)+(-9)(7-9r)=0,$$

or $$98r+98=0,$$

whence $$r=-1.$$

The co-ordinates of the foot of the perpendicular are

$$(-5+r, -3+4r, 6-9r),$$

where $r=-1$ *i.e.*, $(-6, -7, 15)$.

The equations of the line joining $(2, 4, -1)$ and $(-6, -7, 15)$ are

$$\frac{x-2}{8}=\frac{y-4}{11}=\frac{z+1}{-16}, \qquad ...(3)$$

Line (3) is the desired perpendicular from the point $(2, 4, -1)$ on the line (1).

9·4·4. Conditions for Two Lines to Intersect

Given two distinct lines, we have the following possibilities :

(*a*) the lines intersect,

(*b*) the lines do not intersect but are parallel,

(*c*) the lines do not intersect and do not lie in the same plane.

It is often important to know as to which one of the above possibilities holds. (Note that for a pair of distinct lines exactly one of the above possibilities holds).

Case (*b*) is straight-forward and case (*c*) will be taken up a little later. At the moment we are going to discuss case (*a*). Let the equations of two given lines be

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1}=r \qquad ...(i)$$

$$\frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}=\frac{z-z_2}{n_2}=r' \qquad ...(ii)$$

Any point on (*i*) is $(x_1+l_1r, y_1+m_1r, z_1+n_1r)$. Also, any point on (*ii*) is $(x_2+l_2r', y_2+m_2r', z_2+n_2r')$. The lines (*i*) and (*ii*) intersect if and only if there exists a value of r and a value of r' for which the points $(x_1+l_1r, y_1+m_1r, z_1+n_1r)$ and $(x_2+l_2r', y_2+m_2r', z_2+n_2r')$ are the same, *i.e.*, if and only if the system of equations

$$x_1+l_1r=x_2+l_2r',$$
$$y_1+m_1r=y_2+m_2r',$$
$$z_1+m_1r=z_2+n_2r',$$

are consistent,

i.e., if and only if the equations

$$\begin{pmatrix} l_1 & -l_2 \\ m_1 & -m_2 \\ n_1 & -n_2 \end{pmatrix}\begin{pmatrix} r \\ r' \end{pmatrix}=\begin{pmatrix} x_2-x_1 \\ y_2-y_1 \\ z_2-z_1 \end{pmatrix}$$

are consistent.

We shall illustrate the method by examples.

Remarks 1. The lines (*i*) and (*ii*) are parallel if and only if

$$\frac{l_1}{l_2}=\frac{m_1}{m_2}=\frac{n_1}{n_2}.$$

2. In cases (*a*) and (*b*) above, the lines are coplanar. In case (*c*), the lines are non-coplanar.

Example 13. *Show that the lines*

$$\frac{x+1}{3}=\frac{y+3}{5}=\frac{z+5}{7},$$

and $$\frac{x+2}{1}=\frac{y-4}{3}=\frac{z-6}{5}$$

intersect. Find their point of intersection.

Solution. Any point on the line

$$\frac{x+1}{3}=\frac{y+3}{5}=\frac{z+5}{7}=r \text{ (say)} \qquad ...(i)$$

is $(-1+3r, -3+5r, -5+7r)$.

Also any point on the line

$$\frac{x-2}{1}=\frac{y-4}{3}=\frac{z-6}{5}=r' \text{ (say)} \qquad ...(ii)$$

is $(2+r', 4+3r', 6+5r)$.

The lines intersect if and only if

$$\left.\begin{array}{l} -1+3r=2+r' \\ -3+5r=4+3r' \\ -5+7r=6+5r' \end{array}\right\}, \qquad ...(iii)$$

for some values of r and r'.

Equations (*iii*) can be re-written as

$$\left.\begin{array}{l} 3r-r'=3, \\ 5r-3r'=7, \\ 7r-5r'=11. \end{array}\right\} \qquad ...(iv)$$

Solving the first two of the above equations (*iv*), we have $r=\frac{1}{2}$, $r'=-\frac{3}{2}$. Substituting these values of r and r' in the third of the equations (*iv*), we find that they satisfy it. Therefore the system of equations (*iv*) is consistent and $r=\frac{1}{2}, r'=-\frac{3}{2}$ is a common solution. The point of intersection is given by

$$(-1+3r, -3+5r, -5+7r) \text{ where } r=\tfrac{1}{2},$$

i.e., $$\left(\frac{1}{2}, -\frac{1}{2}, \frac{-3}{2}\right).$$

Remark. We could as well have found the point of intersection by substituting

$$r'=-\frac{3}{2} \text{ in } (2+r', 4+3r', 6+5r').$$

9·4·5 Skew Lines. Shortest Distance Between Two Lines

Two straight lines which do not intersect and are not parallel are said to be skew lines. Obviously, two skew lines are never coplanar. Given two skew lines, it is always possible to find a unique line which is perpendicular to both of them. The line segment intercepted by two skew lines on the common perpendicular to both the lines is called the *shortest distance* between the lines.

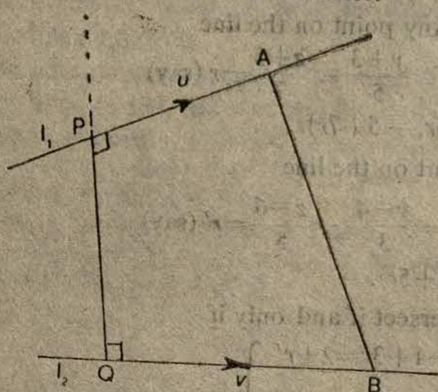


Fig. 9·10.

Let two skew lines l_1 and l_2 be parallel to non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ respectively, and let $\mathbf{a}$ and $\mathbf{b}$ be the position vectors of two points A and B on l_1 and l_2 respectively. The vectors $\mathbf{u}\times\mathbf{v}$ is perpendicular to both l_1 and l_2, and is therefore parallel to their common perpendicular PQ. The shortest distance PQ between the lines is the projection of AB on the common perpendicular. Since the unit vector along the common perpendicular is given by $\frac{\mathbf{u}\times\mathbf{v}}{|\mathbf{u}\times\mathbf{v}|}$, therefore the projection of AB on this vector is given by

$$PQ=\left|(\mathbf{a}-\mathbf{b})\cdot\frac{\mathbf{u}\times\mathbf{v}}{|\mathbf{u}\times\mathbf{v}|}\right|$$
$$=[\mathbf{a}-\mathbf{b},\mathbf{u},\mathbf{v}]/|\mathbf{u}\times\mathbf{v}|.$$

Remark. If the equations of the lines l_1 and l_2 in cartesian form are

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1},$$

and
$$\frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}=\frac{z-z_2}{n_2},$$

then
$$\mathbf{a}=x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k},$$
$$\mathbf{b}=x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k},$$
$$\mathbf{u}=l_1\mathbf{i}+m_1\mathbf{j}+n_1\mathbf{k}$$
$$\mathbf{v}=l_2\mathbf{i}+m_2\mathbf{j}+n_2\mathbf{k},$$

$$[\mathbf{a}-\mathbf{b},\mathbf{u},\mathbf{v}]=\begin{vmatrix} x_1-x_2 & y_1-y_2 & z_1-z_2 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix},$$

$$|\mathbf{u}\times\mathbf{v}|=\sqrt{\sum(m_1n_2-m_2n_1)^2},$$

and therefore the required shortest distance is

$$\text{S.D.}=\frac{\begin{vmatrix} x_1-x_2 & y_1-y_2 & z_1-z_2 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix}}{\sqrt{\sum(m_1n_2-m_2n_1)^2}}.$$

Example 14. *Show that the shortest distance between the lines*

$$\frac{x-1}{2}=\frac{y-2}{3}=\frac{z-3}{4},$$

and
$$\frac{x-2}{3}=\frac{y-4}{4}=\frac{z-5}{5},$$

is $\dfrac{1}{\sqrt{6}}$.

Solution. Here $\mathbf{a}=\mathbf{i}+2\mathbf{j}+3\mathbf{k},$
$$\mathbf{b}=2\mathbf{i}+4\mathbf{j}+5\mathbf{k},$$
$$\mathbf{u}=2\mathbf{i}+3\mathbf{j}+4\mathbf{k},$$
$$\mathbf{v}=3\mathbf{i}+4\mathbf{j}+5\mathbf{k}.$$

$$\therefore\quad \mathbf{u}\times\mathbf{v}=-\mathbf{i}+2\mathbf{j}-\mathbf{k},$$

$$|\mathbf{u}\times\mathbf{v}| = \sqrt{\{(-1)^2+2^2+(-1)^2\}}=\sqrt{6},$$
$$\mathbf{a}-\mathbf{b}=-\mathbf{i}-2\mathbf{j}-2\mathbf{k},$$
$$[\mathbf{a}-\mathbf{b}\ \mathbf{u}\ \mathbf{v}]=(-\mathbf{i}-2\mathbf{j}-2\mathbf{k})\cdot(-\mathbf{i}+2\mathbf{j}-\mathbf{k})=-1.$$
$$\therefore\quad \text{S.D.}=[\mathbf{a}-\mathbf{b}\ \mathbf{u}\ \mathbf{v}]/|\mathbf{u}\times\mathbf{v}|=-1/\sqrt{6}.$$

Since S.D. has to be positive, therefore the required shortest distance $=1/\sqrt{6}$.

Aliter. Here $(x_1, y_1, z_1)=(1, 2, 3)$, $(x_2, y_2, z_2)=(2, 4, 5)$, $(l_1, m_1, n_1)=(2, 3, 4)$, $(l_2, m_2, n_2)=(3, 4, 5)$.

$$\therefore\quad \begin{vmatrix} x_1-x_2 & y_1-y_2 & z_1-z_2 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix} = \begin{vmatrix} -1 & -2 & -2 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{vmatrix}$$
$$=-1\,(3.5-4.4)-(-2).(2.5-3.4)+(-2)\,(2.4-3.3)$$
$$=-1.$$

$$\Sigma\,(m_1n_2-m_2n_1)^2=6.$$

$$\therefore\quad \text{S.D.}=\left|\frac{-1}{\sqrt{6}}\right|=\frac{1}{\sqrt{6}}.$$

We shall now explain another method of finding the shortest distance between two lines. It consists in choosing a point P on the first line and a point Q on the second line such that PQ is perpendicular to both the lines. The additional merit of this method is that it also gives the points where the line of shortest distance meets the given lines, and consequently it enables us to find the equations of the line of shortest distance.

Example 15. *Show that the shortest distance between the lines*

$$\frac{x-3}{3}=\frac{y-8}{-1}=\frac{z-3}{1},$$

and

$$\frac{x-3}{-3}=\frac{y+7}{2}=\frac{z-6}{4}$$

is $3\sqrt{30}$, and that its equations are

$$\frac{x-3}{2}=\frac{y-8}{5}=\frac{z-3}{-1}.$$

Solution. Any point P on the first line is $(3+3r, 8-r, 3+r)$ and any point Q on the second line is $(-3-3r', -7+2r', 6+4r')$. The direction ratios of PQ are

$$(-6-3r-3r', -15+r+2r', 3-r+4r').$$

Let us choose r and r' so that PQ is perpendicular to both the lines. This requires that

$$3(-6-3r-3r')+(-1)(-15+r+2r')+1(3-r+4r')=0,$$

and $$-3(-6-3r-3r')+2(-15+r+2r')+4(3-r+4r')=0.$$

The above equations when simplified yield

$$11r+7r'=0$$
$$7r+29r'=0$$

Solving these equations, we have $r=r'=0$. Therefore P and Q are the points $(3, 8, 3)$ and $(-3, -7, 6)$ respectively.

$$\therefore \quad PQ=\sqrt{\{6^2+15^2+(-3)^2\}}=\sqrt{270}=3\sqrt{30}.$$

The equations of PQ are

$$\frac{x-3}{3-(-3)}=\frac{y-8}{8-(-7)}=\frac{z-6}{3-6},$$

or $$\frac{x-3}{2}=\frac{y-8}{5}=\frac{z-6}{-1}.$$

EXERCISE 9 (*e*)

1. Find the angle between the lines
$$\mathbf{r}=(5\mathbf{i}+2\mathbf{k})+t\,(2\mathbf{j}-5\mathbf{k}),$$
and $$\mathbf{r}=(\mathbf{i}+\mathbf{j})+s\,(3\mathbf{i}+4\mathbf{j}).$$

2. Find the angle between the lines
$$\mathbf{r}=(2\mathbf{i}-\mathbf{j})+s\,(2\mathbf{i}-3\mathbf{k}),$$
and $$\mathbf{r}=(3\mathbf{i}+\mathbf{k})+t\,(\mathbf{i}+\mathbf{j}+\mathbf{k}).$$

3. Show that the lines
$$\mathbf{r}=\mathbf{i}+s\,(\mathbf{i}-\mathbf{j}+\mathbf{k}),$$
and $$\mathbf{r}=\mathbf{j}+t\,(-\mathbf{i}+\mathbf{j}+2\mathbf{k})$$
are perpendicular to each other.

4. Find the angle between the lines
$$\frac{x-1}{2}=\frac{y-2}{3}=\frac{z-3}{4},$$
and $$\frac{x-2}{-3}=\frac{y-3}{4}=\frac{z-3}{4}.$$

5. Find the angle between the lines
$$\frac{x+1}{1}=\frac{y-2}{-2}=\frac{z-4}{2},$$
and $$\frac{x-1}{2}=\frac{y-3}{2}=\frac{z-1}{1}.$$

6. Find the distance between the lines

$$\frac{x}{-1}=\frac{y-2}{3}=\frac{z+1}{2},$$

and $$\frac{x-1}{2}=\frac{y}{-6}=\frac{z+3}{-4}.$$

Are the lines coplanar ?

7. Show that the lines

$$\frac{x-1}{2}=\frac{y-2}{3}=\frac{z-3}{4},$$

and $$\frac{x-2}{3}=\frac{y-3}{4}=\frac{z-4}{5}$$

intersect.

8. Show that the lines

$$\frac{x-5}{4}=\frac{y-7}{4}=\frac{z+3}{-5},$$

and $$\frac{x-8}{7}=\frac{y-4}{1}=\frac{z-5}{3}$$

intersect, and find the point of intersection.

9. Show that the lines

$$\frac{x}{1}=\frac{y-2}{2}=\frac{z+3}{3},$$

and $$\frac{x-2}{2}=\frac{y-6}{3}=\frac{z-3}{4}$$

intersect, and find their point of intersection.

10. Find the equations of the perpendicular drawn from the point $(3, -1, 11)$ to the line

$$\frac{x}{2}=\frac{y-2}{3}=\frac{z-3}{4}.$$

11. Find the length of the shortest distance between the lines

$$\frac{x-3}{1}=\frac{y-5}{-2}=\frac{z-7}{7}.$$

$$\frac{x+1}{7}=\frac{y+1}{-6}=\frac{z+1}{1}.$$

Also find its equations.

12. Find the co-ordinates of the points where the line of shortest distance between the lines

$$\frac{x-12}{-9}=\frac{y-1}{4}=\frac{z-5}{2},$$

and $$\frac{x-23}{-6}=\frac{y-19}{-4}=\frac{z-25}{3}$$

meets them.

13. Find the magnitude and equations of the line of shortest distance between the lines

$$\frac{x-1}{1}=\frac{y+7}{3}=\frac{z+2}{2},$$

and $$\frac{x-3}{-1}=\frac{y-4}{2}=\frac{z+2}{1}.$$

14. Find the length and equations of the line of shortest distance between the lines

$$\frac{x+1}{2}=\frac{y-1}{1}=\frac{z-9}{-3},$$

and $$\frac{x-3}{2}=\frac{y+15}{-7}=\frac{z-9}{5}.$$

15. Find the magnitude and the equations of the line of shortest distance between the lines

$$\frac{x-8}{3}=\frac{y+9}{-16}=\frac{z-10}{7},$$

and $$\frac{x-15}{3}=\frac{y-29}{8}=\frac{z-5}{-5}.$$

Also find the co-ordinates of the points where the line of shortest distance meets the given lines.

9·5. THE PLANE

We shall now study planes. We shall first of all determine the vector equation of a plane in different forms and translate the same into cartesian form. We shall also find the angle between two planes, the angle between a line and a plane, and the length of the perpendicular from a given point on a plane. We shall study the family of planes passing through the line of intersection of two planes.

9·5·1. Equation of a Plane in Terms of its Distance From the Origin and Normal to a given Direction

Let **n** be the unit vector normal to a given plane, and let p be the length of the perpendicular from the origin of reference to the plane. We shall always consider p to be positive.

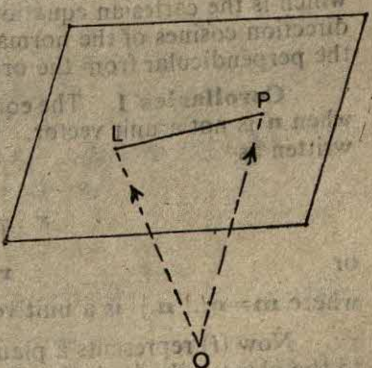


Fig. 9·11.

Draw OL perpendicular to the plane, L being the foot of the perpendicular. Since **n** is the unit vector normal to the plane, and p is the length of the perpendicular from O on the plane, therefore

$$\overrightarrow{OL}=p\mathbf{n}.$$

Let P be any point on the plane and let $\mathbf{r}$ be the position vector of P, so that

$$\overrightarrow{OP}=\mathbf{r}.$$

Now LP lies in the plane and $\mathbf{n}$ is normal to the plane.

Therefore $\quad \overrightarrow{LP} \cdot \mathbf{n}=0,$

i.e., $\quad \overrightarrow{OP}-\overrightarrow{OL} \cdot \mathbf{n}=0,$

or $\quad (\mathbf{r}-p\mathbf{n}) \cdot \mathbf{n}=0,$

or $\mathbf{r} \cdot \mathbf{n}=p\,\mathbf{n} \cdot \mathbf{n}=p$, since $\mathbf{n}$ is a unit vector. Thus we find that the equation

$$\boxed{\mathbf{r} \cdot \mathbf{n}=p} \qquad \text{...(A)}$$

is satisfied by the position vector of every point on the plane.

Also, conversely, any point P whose position vector $\mathbf{r}$ satisfies (A), is a point on the plane.

Thus $\mathbf{r} \cdot \mathbf{n}=p$ is the vector equation of the plane, where $\mathbf{n}$ is the unit vector normal to the plane, and p is the length of the perpendicular from the origin to the plane.

If $\mathbf{n}=l\mathbf{i}+m\mathbf{j}+n\mathbf{k}$, then (A) can be written as

$$(x\mathbf{i}+y\mathbf{j}+z\mathbf{k}) \cdot (l\mathbf{i}+m\mathbf{j}+n\mathbf{k})=p,$$

or

$$\boxed{l\,x+my+nz=p} \qquad \text{...(B)}$$

which is the cartesian equation of a plane, where (l, m, n) are the direction cosines of the normal to the plane, and p is the length of the perpendicular from the origin to the plane.

Corollaries 1. The equation $\mathbf{r} \cdot \mathbf{n}=q$ represents a plane even when $\mathbf{n}$ is not a unit vector. For, in that case $\mathbf{r} \cdot \mathbf{n}=q$ can be re-written as

$$\mathbf{r} \cdot \frac{\mathbf{n}}{|\mathbf{n}|}=\frac{q}{|\mathbf{n}|},$$

or

$$\mathbf{r} \cdot \mathbf{m}=\mathrm{p}, \qquad \text{...}(i)$$

where $\mathbf{m}=\mathbf{n}/|\mathbf{n}|$ is a unit vector, and $p=q/|\mathbf{n}|$.

Now (i) represents a plane such that $\mathbf{m}$ is a unit vector normal to the plane and p is the length of the perpendicular from the origin

on the plane. Thus we find that $\mathbf{r}\cdot\mathbf{n}=q$ *represents a plane such that* $\mathbf{n}$ *is a vector normal to the plane and* $q/|\mathbf{n}|$ *is the length of the perpendicular from the origin on the plane.*

2. The equation $ax+by+cz+d=0$ represents a plane such that (a, b, c) are the direction ratios of the normal to the plane and $\dfrac{|d|}{\sqrt{(a^2+b^2+c^2)}}$ is the length of the perpendicular from the origin on the plane.

This can be seen by re-writing the equation in the form

$$(a\mathbf{i}+b\mathbf{j}+c\mathbf{k})\cdot(x\mathbf{i}+y\mathbf{j}+z\mathbf{k})=-d.$$

9.5.2. Equation of a Plane Containing a given Point A with Position Vector $\mathbf{r}_1$ and normal to given Vector $\mathbf{n}$. Let $\mathbf{r}$ be the position vector of any point P on the plane.

$$\vec{AP}=\vec{OP}-\vec{OA}=\mathbf{r}-\mathbf{r}_1.$$

Since $\vec{AP}$ lies in the plane and $\mathbf{n}$ is normal to the plane, therefore

$$(\mathbf{r}-\mathbf{r}_1)\cdot\mathbf{n}=0$$

Thus we find that *the vectorial equation of the plane containing the point* $\mathbf{r}_1$ *and normal to the vector* $\mathbf{n}$ *is*

$$\boxed{(\mathbf{r}-\mathbf{r}_1)\cdot\mathbf{n}=0} \qquad \ldots(A)$$

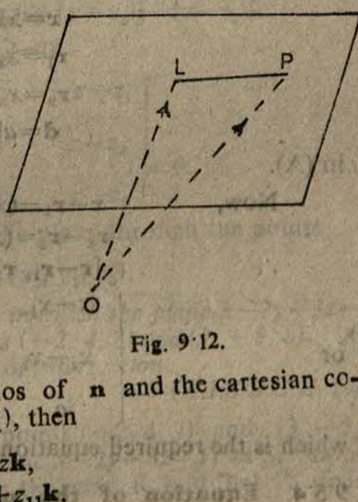


Fig. 9·12.

If (a, b, c) be the direction ratios of $\mathbf{n}$ and the cartesian co-ordinates of the point A be (x_1, y_1, z_1), then

$$\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k},$$
$$\mathbf{r}_1=x_1\mathbf{i}+y_1\mathbf{j}+z_{11}\mathbf{k},$$
$$\mathbf{n}=a\mathbf{i}+b\mathbf{j}+c\mathbf{k},$$

so that $(\mathbf{r}-\mathbf{r}_1)\cdot\mathbf{n}=[(x-x_1)\mathbf{i}+(y-y_1)\mathbf{j}+(z-z_1)\mathbf{k}]\cdot[a\mathbf{i}+b\mathbf{j}+c\mathbf{k}]$
$$=a(x-x_1)+b(y-y_1)+c(z-z_1)$$

(A) can therefore be re-written as

$$\boxed{a(x-x_1)+b(y-y_1)+c(z-z_1)=0} \qquad \ldots(B)$$

(B) represents a plane passing through the point (x_1, y_1, z_1) and normal to the line having direction-ratios (a, b, c).

9·5·3. Equation of the Plane Through two given Points A, B with Position Vectors $\mathbf{r}_1$, $\mathbf{r}_2$ and parallel to a given Vector d

Since the points A, B lie in the plane, therefore the vector $\overrightarrow{AB}$ is parallel to the plane, *i.e.*, $\mathbf{r}_2-\mathbf{r}_1$ is parallel to the plane. The vectors $\mathbf{r}_2-\mathbf{r}_1$ and $\mathbf{d}$ being both parallel to the plane, $(\mathbf{r}_2-\mathbf{r}_1)\times\mathbf{d}$ is normal to the plane. We have to find the equation of the plane containing the point with position vector $\mathbf{r}_1$ and normal to the vector $(\mathbf{r}_2-\mathbf{r}_1)\times\mathbf{d}$. As in 9·5·2 above, the equation of the plane is

$$\boxed{(\mathbf{r}-\mathbf{r}_1)\cdot[(\mathbf{r}_2-\mathbf{r}_1)\times\mathbf{d}]=0} \qquad \text{...(A)}$$

Corollary. To find the cartesian equation of the plane passing through the points (x_1, y_1, z_1) and (x_2, y_2, z_2) and parallel to a line having direction ratios (a, b, c), we have to substitute

$$\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k},$$
$$\mathbf{r}_1=x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k},$$
$$\mathbf{r}_2=x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k},$$
$$\mathbf{d}=a\mathbf{i}+b\mathbf{j}+c\mathbf{k},$$

in (A).

Now, $$\mathbf{r}-\mathbf{r}_1=(x-x_1)\mathbf{i}+(y-y_1)\mathbf{j}+(z-z_1)\mathbf{k},$$
$$\mathbf{r}_2-\mathbf{r}_1=(x_2-x_1)\mathbf{i}+(y_2-y_1)\mathbf{j}+(z_2-z_1)\mathbf{k}.$$

$\therefore$ $$[\mathbf{r}-\mathbf{r}_1, \mathbf{r}_2-\mathbf{r}_1, \mathbf{d}]=0,$$

or $$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ a & b & c \end{vmatrix}=0, \qquad \text{...(B)}$$

which is the required equation.

9·5·4. Equation of the plane Through Three Points A, B, C with Position Vectors $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$.

Let A, B, C be three points with position vectors $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ respectively and let $\mathbf{r}$ be the position vector of any point P in the plane.

The vectors $\overrightarrow{AP}$, $\overrightarrow{BP}$, $\overrightarrow{CP}$ are coplanar, and therefore

$$[\overrightarrow{AP}, \overrightarrow{BP}, \overrightarrow{CP}]=0.$$

But $$\overrightarrow{AP}=\mathbf{r}-\mathbf{r}_1,\ \overrightarrow{BP}=\mathbf{r}-\mathbf{r}_2,\ \overrightarrow{CP}=\mathbf{r}-\mathbf{r}_3.$$

$$\therefore \quad [\overrightarrow{AP}, \overrightarrow{BP}, \overrightarrow{CP}]=0$$

$$\Leftrightarrow [\mathbf{r}-\mathbf{r}_1 \ \mathbf{r}-\mathbf{r}_2 \ \mathbf{r}-\mathbf{r}_3]=0, \qquad \text{...(A)}$$

which is the required equation.

Corollary. To find the cartesian equation of the plane passing through the points (x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3) we have to substitute

$$\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}.$$
$$\mathbf{r}_1=x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k},$$
$$\mathbf{r}_2=x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k},$$
$$\mathbf{r}_3=x_3\mathbf{i}+y_3\mathbf{j}+z_3\mathbf{k},$$

in (A).

Now
$$\mathbf{r}-\mathbf{r}_1=(x-x_1)\mathbf{i}+(y-y_1)\mathbf{j}+(z-z_2)\mathbf{k},$$
$$\mathbf{r}-\mathbf{r}_2=(x-x_2)\mathbf{i}+(y-y_2)\mathbf{j}+(z-z_2)\mathbf{k},$$
$$\mathbf{r}-\mathbf{r}_3=(x-x_3)\mathbf{i}+(y-y_3)\mathbf{j}+(z-z_3)\mathbf{k}.$$

$$[\mathbf{r}-\mathbf{r}_1 \ \mathbf{r}-\mathbf{r}_2 \ \mathbf{r}-\mathbf{r}_3]=0$$

$$\Leftrightarrow \begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x-x_2 & y-y_2 & z-z_2 \\ x-x_3 & y-y_3 & z-z_3 \end{vmatrix}=0, \qquad \text{...(B)}$$

which is the required equation of the plane through the points

$$(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3).$$

Example 16. *Find the ratio in which the plane $x-2y+3z=17$ divides the line joining the points $(-2, 4, 7)$ and $(3, -5, 8)$. Also obtain the co-ordinates of the point of intersection.* (*A.I.S.S.C.E. 1988*)

Solution. Let us name the points $(-2, 4, 7)$ and $(3, -5, 8)$ as P and Q respectively, and the plane $x-2y+3z=17$ as S. Suppose that the join of P and Q meets S in the point R, and that R divides PQ in the ratio $k : 1$.

The co-ordinates of R are

$$\left(\frac{3k-2}{k+1}, \frac{-5k+4}{k+1}, \frac{8k+7}{k+1}\right). \qquad \text{...(1)}$$

R lies on S provided

$$\frac{3k-2}{k+1}-2.\frac{-5k+4}{k+1}+3.\frac{8k+7}{k+1}=17,$$

or $\quad (3k-2)-2(-5k+4)+3(8k+7)-17(k+1)=0,$

or $\quad 20k-6=0,$

so that $$k=\frac{3}{10}.$$

$\therefore$ The required ratio $=3:10$.

The co-ordinates of the point of intersection are obtained by putting $k=\frac{3}{10}$ in (1). Putting $k=\frac{3}{10}$ in (1) we get the co-ordinates of R as $\left(-\frac{11}{13}, \frac{25}{13}, \frac{94}{13}\right)$.

Example 17. *Find the equation of the plane through the points $(-2, -2, 2)$, $(1, 1, 1)$, and $(1, -1, 2)$.*

Solution. The equation of any plane through the point $(-2, -2, 2)$ is

$$A(x+2)+B(y+2)+C(z-2)=0. \qquad \ldots(1)$$

It passes through the points $(1, 1, 1)$ and $(1, -1, 2)$ provided

$$3A+3B-C=0, \qquad \ldots(2)$$

$$3A+B\quad=0, \qquad \ldots(3)$$

Eliminating A, B, C from equations (1), (2) and (3), we have

$$\begin{vmatrix} x+2 & y+2 & z-2 \\ 3 & 3 & -1 \\ 3 & 1 & 0 \end{vmatrix}=0, \qquad \ldots(4)$$

as the desired equation.

By expanding the determinant on the left hand side of (4), we can re-write (4) as

$$x-3y-6z+8=0.$$

EXERCISE 9 (*f*)

1. Write the vectorial equation of the plane having $\mathbf{j}$ as the unit vector normal to it, and at a distance 2 units from the origin of reference.
2. A plane is normal to the vector $\mathbf{i}+2\mathbf{j}-2\mathbf{k}$. If its distance from the origin be 5 units, find its vectorial equation.
3. A plane is normal to the vector $2\mathbf{i}-\mathbf{j}+2\mathbf{k}$. If its equation is $\mathbf{r}.(2\mathbf{i}-\mathbf{j}+2\mathbf{k})=6$, find its distance from the origin.
4. Find the equation to the plane through $(1, 2, 3)$ parallel to the plane $3x+4y-5z=0$.
5. Find the equation of the plane which is at distance 3 units from the origin and has $(-1, 2, 2)$ as the direction ratios of a normal to it.

6. Find the equation of the plane passing through the points with position vectors 3**i** and 2**j** ond having **j**−**k** as a normal to it.
7. Find the equation of the plane passing through the points (1, 0, 1), (1, 1, 0) and (0, 1, 1).
8. Find the vectorial equation of the plane passing through the points with position vectors **i**−**j** and **j**+**k**, and parallel to the vector **i**−**k**. Find the cartesian equation also.
9. Find the equation of the plane through the points A (7, 0, 6), B (3, 4, 2), and C (2, 2, −1).
10. Find the equation to the plane through the points (−7, −3, −5), (1, −1, 1) and (1, 1, 1).
11. Find the equation to the plane through the points (2, 3, 1) and (4, −5, 3), and parallel to the x-axis.
12. Find the equation to the plane through the points (0, −1, 0), (2, 1, −1) and (1, 1, 1).
13. Show that the four points (0, −1, 0), (2, 1, −1), (1, 1, 1) and (3, 3, 0) are coplanar.

9·5·5. Angle Between two Planes

The angle between two given planes is equal to the angle between the vectors normal to the planes.

Suppose that the equations of two planes are

$$\mathbf{r} \cdot \mathbf{n}_1 = q_1 \text{ and } \mathbf{r} \cdot \mathbf{n}_2 = q_2, \qquad ...(1)$$

where $\mathbf{n}_1$ and $\mathbf{n}_2$ are vectors normal to the two planes. The angle θ between the planes is given by

$$\boxed{\cos\theta = \frac{\mathbf{n}_1 \cdot \mathbf{n}_2}{|\mathbf{n}_1|\ |\mathbf{n}_2|}.} \qquad ...(A)$$

If the equations are given in cartesian form, we can find the angle between them as follows :

Let the equations of two planes be

$$S_1 \equiv a_1x + b_1y + c_1z + d_1 = 0, \qquad ...(2)$$

$$S_2 \equiv a_2x + b_2y + c_2z + d_2 = 0. \qquad ...(3)$$

The direction ratios of the normals to S_1 and S_2 are (a_1, b_1, c_1) and (a_2, b_2, c_2) respectively. Therefore the angle between the planes is given by

$$\boxed{\cos\theta = \frac{a_1a_2 + b_1b_2 + c_1c_2}{\sqrt{(a_1^2+b_1^2+c_1^2)}\ \sqrt{(a_2^2+b_2^2+c_2^2)}}} \qquad ...(B)$$

Example 18. *Find the equation of the plane through the points (2, 2, 1) and (9, 3, 6), and perpendicular to the plane $2x+6y+6z=9$.*
(*A.I.S.S.C.E.*, *1984*)

Solution. Since the required plane is perpendicular to the plane $2x+6y+6z=9$, therefore it is parallel to the line having direction ratios (2, 6, 6), *i.e.*, it is parallel to the vector $2\mathbf{i}+6\mathbf{j}+6\mathbf{k}$.

Since the points with position vectors $2\mathbf{i}+2\mathbf{j}+\mathbf{k}$ and $9\mathbf{i}+3\mathbf{j}+6\mathbf{k}$ lie in the plane, the vector $(2\mathbf{i}+2\mathbf{j}+\mathbf{k})-(9\mathbf{i}+3\mathbf{j}+6\mathbf{k})$ is parallel to the plane, *i.e.*, the vector $-7\mathbf{i}-\mathbf{j}-5\mathbf{k}$ is parallel to the plane. We have to find the equation of the plane passing through the point with position vector $2\mathbf{i}+2\mathbf{j}+\mathbf{k}$, and parallel to the vectors $2\mathbf{i}+6\mathbf{j}+6\mathbf{k}$ and $-7\mathbf{i}-\mathbf{j}-5\mathbf{k}$. The required equation is

$$[\mathbf{r}-(2\mathbf{i}+2\mathbf{j}+\mathbf{k})\ 2\mathbf{i}+6\mathbf{j}+6\mathbf{k}\ -7\mathbf{i}-\mathbf{j}-5\mathbf{k}]=0, \quad \ldots(1)$$

i.e.
$$\begin{vmatrix} x-2 & y-2 & z-1 \\ 2 & 6 & 6 \\ -7 & -1 & -5 \end{vmatrix}=0,$$

or $$(x-2)(-24)-(y-2)(32)+(z-1)(40)=0,$$

or $$3x+4y-5z-9=0.$$

Aliter. The equation of any plane through the point (2, 2, 1) is

$$A(x-2)+B(y-2)+C(z-1)=0. \quad \ldots(1)$$

(1) passes through the point (9, 3, 6) provided

$$A(9-2)+B(3-2)+C(6-1)=0,$$

i.e., $$7A+B+5C=0. \quad \ldots(2)$$

Also, the plane (1) is perpendicular to the plane

$$2x+6y+6z=9 \text{ provided}$$
$$2A+6B+6C=0. \quad \ldots(3)$$

Eliminating A, B, C from (1), (2), and (3) we have

$$\begin{vmatrix} x-2 & y-2 & z-1 \\ 7 & 1 & 5 \\ 2 & 6 & 6 \end{vmatrix}=0,$$

i.e., $$3x+4y-5z-9=0.$$

9·5·6. Angle Between a Line and a Plane

Let $$\mathbf{r}=\mathbf{r}_1+t\mathbf{b}_1, \quad \ldots(1)$$
be the vector equation of a line and let
$$\mathbf{r}\cdot\mathbf{n}=q \quad \ldots(2)$$
be the vector equation of a plane.

The angle between (1) and (2) is the complement of the angle between the direction vector **b** of the given line and the normal vector **n** of the given plane. If θ be the angle between (1) and (2), then

$$\cos(\pi/2-\theta)=\frac{\mathbf{n}\cdot\mathbf{b}}{|\mathbf{n}|\,|\mathbf{b}|},$$

or

$$\boxed{\sin\theta=\frac{\mathbf{n}\cdot\mathbf{b}}{|\mathbf{n}|\,|\mathbf{b}|}.} \qquad \ldots\text{A)}$$

If the equations of the line be

$$\frac{x-x_1}{l}=\frac{y-y_1}{m}=\frac{z-z_1}{n}, \qquad \ldots(3)$$

and that of the plane be

$$ax+by+cz+d=0, \qquad \ldots(4)$$

then
$$\mathbf{b}=l\mathbf{i}+m\mathbf{j}+n\mathbf{k},$$
$$\mathbf{n}=a\mathbf{i}+b\mathbf{j}+c\mathbf{k},$$

so that form (A) we get

$$\boxed{\sin\theta=\frac{al+bm+cn}{\sqrt{(a^2+b^2+c^2)}\sqrt{(l^2+m^2+n^2)}}} \qquad \ldots(\text{B})$$

Corollary. If the line (3) is parallel to the plane (4), then $\theta=0$, and consequently (B) gives

$$al+bm+cn=0.$$

Thus we conclude that if the line

$$\frac{x-x_1}{l}=\frac{y-y_1}{m}=\frac{z-z_1}{n}$$

is parallel to the plane

$$ax+by+cz+d=0,$$

then
$$al+bm+cn=0.$$

Example 19. *Find the equation of the plane containing the line*

$$\frac{x+2}{2}=\frac{y+3}{3}=\frac{z-4}{-2}$$

and the point (0, 6, 0).

Solution. The equation of any plane through the point (0, 6, 0) is

$$\mathrm{A}x+\mathrm{B}(y-6)+\mathrm{C}z=0 \qquad \ldots(1)$$

If (1) contains the line

$$\frac{x+2}{2}=\frac{y+3}{3}=\frac{z-4}{-2}, \qquad ...(2)$$

it must pass through the point $(-2, -3, 4)$ and must be parallel to the line (2).

Since (1) must pass through the point $(-2, -3, 4)$,

therefore $\quad A(-2)+B(-9)+C(4)=0,$

i.e., $$2A \quad +9B \quad +\ 4C=0 \qquad ...(3)$$

Also (1) is parallel to the line (2) provided

$$2A \quad +3B \quad -2C=0. \qquad ...(4)$$

Eliminating A, B, C from (1), (3) and (4), we have

$$\begin{vmatrix} x & y-6 & z \\ 2 & 9 & -4 \\ 2 & 3 & -2 \end{vmatrix}=0,$$

or $$x[9(-2)-(-4).3]-(y-6)[2(-2)-(-4).2]$$
$$+z(2.3-2.9)=0,$$

or $$3x+2y+6z-12=0.$$

Example 20. *Find the angle between the planes $2x-y+z=6$ and $x+y+2z=7$.*

Solution. The direction ratios of the normals to the two planes are given by $(2, -1, 1)$ and $(1, 1, 2)$. The angle θ between the planes is given by

$$\cos\theta=\frac{2.1+(-1).1+1.2}{\sqrt{(2^2+(-1)^2+1^2)}\sqrt{(1^2+1^2+2^2)}},$$
$$=\frac{1}{2}.$$

$\therefore \qquad \theta=\pi/3.$

Thus the planes are inclined to each other at an angle $\pi/3$.

9·5·7. Distance of a Point From a Plane.

Let $\mathbf{r}_1$ be the position vector of a given point A and

$$\mathbf{r} \cdot \mathbf{n}=q \qquad ...(1)$$

the equation of the given plane. The equation of the line through A, normal to the plane is

$$\mathbf{r}=\mathbf{r}_1+t\mathbf{n} \qquad ...(2)$$

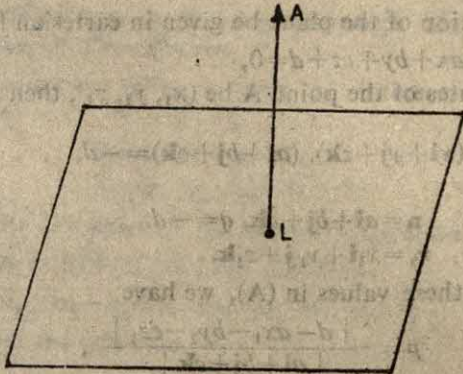


Fig. 9·13.

For the point of intersection of (1) and (2), t is obtained by solving (1) and (2) together. Substituting the value of $\mathbf{r}$ from (2) in (1), we have

$$(\mathbf{r}_1+t\mathbf{n}).\mathbf{n}=q,$$

or $$\mathbf{r}_1.\mathbf{n}+t\mathbf{n}.\mathbf{n}=q,$$

so that $$t=\frac{q-\mathbf{r}_1 \,.\, \mathbf{n}}{\mathbf{n} \,.\, \mathbf{n}} \qquad ...(3)$$

From (2) and (3) we find that the position vector of L, the foot of the perpendicular from A on the given plane, is given by

$$\mathbf{r}=\mathbf{r}_1+\frac{q-\mathbf{r}_1 \,.\, \mathbf{n}}{\mathbf{n} \,.\, \mathbf{n}}\mathbf{n} \qquad ...(4)$$

$$\therefore \qquad \text{AL}=|\overrightarrow{\text{AL}}|$$
$$=|\mathbf{r}-\mathbf{r}_1|$$
$$=\left|\frac{q-\mathbf{r}_1 \cdot \mathbf{n}}{\mathbf{n} \,.\, \mathbf{n}}\mathbf{n}\right|$$
$$=\frac{|q-\mathbf{r}_1 \,.\, \mathbf{n}|}{|\mathbf{n}|^2}.|\mathbf{n}|$$
$$=\frac{|q-\mathbf{r}_1 \,.\, \mathbf{n}|}{|\mathbf{n}|}.$$

Thus we find that *the length of the perpedicular from the point* $\mathbf{r}_1$ *on the plane* $\mathbf{r} \,.\, \mathbf{n}=q$ *is given by*

$$\boxed{p=\frac{|q-\mathbf{r}_1 \,.\, \mathbf{n}|}{\mathbf{n}}} \qquad ...(A)$$

If the equation of the plane be given in cartesian form as

$$ax+by+cz+d=0, \qquad \ldots(5)$$

and the co-ordinates of the point A be (x_1, y_1, z_1), then rewriting (5) as

$$(x\mathbf{i}+y\mathbf{j}+z\mathbf{k}).(a\mathbf{i}+b\mathbf{j}+c\mathbf{k})=-d,$$

we find that

$$\mathbf{n}=a\mathbf{i}+b\mathbf{j}+c\mathbf{k},\ q=-d.$$

Also, $\quad \mathbf{r}_1=x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k}.$

Substitutin these values in (A), we have

$$p=\frac{|\,d-ax_1-by_1-cz_1\,|}{|\,a\mathbf{i}+b\mathbf{j}+c\mathbf{k}\,|},$$

i.e.,

$$\boxed{p=\frac{|\,ax_1+by_1+cz_1+d\,|}{\sqrt{(a^2+b^2+c^2)}}} \qquad \ldots(B)$$

Example 21. *Find the length and the foot of the perpendicular from the point (7, 14, 5) to the plane $2x+4y-z=2$.*

(A.I.S.S.C.E., 1987)

Solution. The equations of the line through (7, 14, 5) perpendicular to the plane

$$2x+4y-z=2 \qquad \ldots(i)$$

are

$$\frac{x-7}{2}=\frac{y-14}{4}=\frac{z-5}{-1}=r \text{ (say)} \qquad \ldots(ii)$$

Any point P on (*ii*) is $(2r+7, 4r+14, -r+5)$. P is the foot of the perpendicular from the given point on (*i*) provided the co-ordinates of P satisfy (*i*).

$$\therefore \quad 2(2r+7)+4(4r+14)-(-r+5)=2,$$

or $\quad 21r=-63,$

so that $\quad r=-3.$

Substituting $r=-3$ in $(2r+7, 4r+14, -r+5)$ we find that the co-ordinates of the foot of the perpendicular are (1, 2, 8). Also, the length of the perpendicular, p, is equal to the distance between the points (7, 14, 5) and (1, 2, 8).

$$p=\sqrt{\{(7-1)^2+(14-2)^2+(5-8)^2\}}$$

$$=\sqrt{189}=3\sqrt{21}$$

Verification. Using formula (B), we have

$$p=\frac{|\,7.2+14.4+5(-1)-21\,|}{\sqrt{\{2^2+4^2+(-1)^2\}}}$$

$$=\frac{63}{\sqrt{21}}=3\sqrt{21}.$$

9·5·8. Family of Planes Passing Through the Line of Intersection of Two Planes

Suppose we are given a pair of intersecting planes whose equations are

$$S_1 = \mathbf{r} \cdot \mathbf{n}_1 - q_1 = 0, \quad \ldots(i)$$
$$S_2 = \mathbf{r} \cdot \mathbf{n}_2 - q_2 = 0. \quad \ldots(ii)$$

Let $\mathbf{r}_1$ be any point on the line of intersection of (*i*) and (*ii*).

Since $\mathbf{r}_1$ satisfies (*i*) and (*ii*), therefore

$$\mathbf{r}_1 \cdot \mathbf{n}_1 - q_1 = 0, \text{ and } \mathbf{r}_1 \cdot \mathbf{n}_2 - q_2 = 0 \quad \ldots(iii)$$

Consider the equation

$$p\,(\mathbf{r} \cdot \mathbf{n}_1 - q_1) + (\mathbf{r} \cdot \mathbf{n}_2 - q_2) = 0 \quad \ldots(iv)$$

For all values of p and q, (*iv*) represents a plane. Also, because of (*iii*), we have

$$p\,(\mathbf{r}_1 \cdot \mathbf{n}_1 - q_1) + q(\mathbf{r}_1 \cdot \mathbf{n}_2 - q_2) = 0,$$

showing that (*iv*) passes through $\mathbf{r}_1$ which is a point on the line of intersection of the planes (*i*) and (*ii*).

The equation $pS_1 + qS_2 = 0$ represents the family of all planes passing through the line of intersection of the planes $S_1 = 0$ and $S_2 = 0$. It appears as if two *independent* parameters p and q are involved here. But this is not the case. Observe that p and q cannot be both zero. If we impose one condition (other than that of passing through the line of intersection of $S_1 = 0$ and $S_2 = 0$) such as passing through a given point, or being perpendicular to a given plane, then the ratio $p : q$ can be determined, which is exactly what we require.

It may be noted that $S_1 + kS_2 = 0$ represents the family of all planes other than $S_2 = 0$ that pass through the line of intersection of $S_1 = 0$ and $S_2 = 0$. Similarly $S_2 + k'S_1 = 0$ represents the family of all planes other than $S_1 = 0$ that pass through the line of intersection of $S_1 = 0$ and $S_2 = 0$.

Remark. If the equation of the planes are given in cartesian co-ordinates as

$$S_1 \equiv a_1x + b_1y + c_1z + d_1 = 0,$$
$$S_2 \equiv a_2x + b_2y + c_2z + d_2 = 0,$$

the above discussion holds with obvious modifications and the equation $pS_1 + qS_2 = 0$ represents the family of all planes passing through the line of intersection of the planes

$$S_1 = 0 \text{ and } S_2 = 0.$$

Example 22. *Find the equation of the plane passing through the intersection of the planes $2x + 3y + 4z - 5 = 0$ and $x - y + z - 1 = 0$, and passing through the point $(3, -2, 1)$.*

Solution. The equations of the given planes are

$$2x+3y+4z-5=0, \quad \text{...}(i)$$

$$x-y+z-1=0. \quad \text{...}(ii)$$

The equation of any plane [except (*ii*)] through the line of intersection of (*i*) and (*ii*) is

$$(2x+3y+4z-5)+k(x-y+z-1)=0. \quad \text{...}(iii)$$

If (*iii*) passes through the point (3, −2, 1), then

$$[3.2+3(-2)+4.1-5]+k[3-(-2)+1-1]=0,$$

or $$-1+5k=0,$$

so that $$k=1/5.$$

Putting $k=1/5$ in (*iii*), we have

$$(2x+3y+4z-5)+\tfrac{1}{5}(x-y+z-1)=0,$$

or $$11x+14y+21z-26=0,$$

as the required equation.

Example 23. *Find the equation of the plane which contains the line of intersection of the planes*

$$x+2y+3z-4=0 \text{ and } 2x+y-z+5=0$$

and which is perpendicular to the plane

$$5x+3y-6z+8=0. \qquad (D.B.S.S.C.E., 1987)$$

Solution. The equations of the given planes are

$$x+2y+3z-4=0, \quad \text{...}(i)$$

$$2x+y-z+5=0. \quad \text{...}(ii)$$

The equation of any plane except (*ii*) through the line of intersection of (*i*) and (*ii*) is

$$(x+2y+3z-4)+k(2x+y-z+5)=0, \quad \text{...}(iii)$$

or $$(1+2k)\,x+(2+k)\,y+(3-k)z-4+5k=0, \quad \text{...}(iv)$$

If (*iv*) is pespendicular to $5x+3y-6z+8=0$, then

$$5(1+2k)+3(2+k)-6(3-k)=0,$$

or $$k=7/19.$$

Putting $k=7/19$ in (*iii*), we have

$$(x+2y+3z-4)+(7/19)(2x+y-z+5)=0,$$

or $$33x+45y+50z-41=0,$$

as the required equation.

Example 24. *Find the equations of the plane passing through the line of intersection of the planes* $\mathbf{r}\cdot\mathbf{n}_1=q_1$ *and* $\mathbf{r}\cdot\mathbf{n}_2=q_2$ *and containing the point* $\mathbf{a}$.

Solution. Let $\mathbf{r}\cdot\mathbf{n}_1=q_1$ and $\mathbf{r}\cdot\mathbf{n}_2=q_2$ ...(1)
be the equations of two planes.

The equation

$$(\mathbf{r}\cdot\mathbf{n}_1-q_1)+k(\mathbf{r}\cdot\mathbf{n}_2-q_2)=0 \qquad ...(2)$$

represents a plane passing through the line of intersection of the given planes, whatever the value of k may be. The plane (2) passes through the point $\mathbf{a}$ provided

$$(\mathbf{a}\cdot\mathbf{n}_1-q_1)+k(\mathbf{a}\cdot\mathbf{n}_2-q_2)=0,$$

or
$$k=-(\mathbf{a}\cdot\mathbf{n}_1-q_1)/(\mathbf{a}\cdot\mathbf{n}_2-q_2). \qquad ...(3)$$

Substituting the value of k in (2), we find that the required equation is

$$(\mathbf{r}\cdot\mathbf{n}_1-q_1)-[(\mathbf{a}\cdot\mathbf{n}_1-q_1)/(\mathbf{a}\cdot\mathbf{n}_2-q_2)],\ (\mathbf{r}\mathbf{n}_2-\mathbf{q}_2)]=0,$$

or
$$(\mathbf{r}\cdot\mathbf{n}_1-q_1)(\mathbf{a}\cdot\mathbf{n}_2-q_2)-(\mathbf{a}\cdot\mathbf{n}_1-q_1)(\mathbf{r}\cdot\mathbf{n}_2-q_2)=0.$$

EXERCISE 9 (*g*)

1. Find the angle between the planes
$3x+4y-5z=9$ and $2x+6y+6z=7$.

2. Find the angle between the planes $\mathbf{r}\cdot(\mathbf{i}-\mathbf{j})=1$ and $\mathbf{r}\cdot(\mathbf{j}+\mathbf{k})=2$.

3. Find tee angle between the line $\dfrac{x}{3}=\dfrac{y}{2}=\dfrac{z}{4}$ and the plane
$2x+3y+4z-18=0$.

4. Find the length of the perpendicular drawn from the origin to the plane $6x-3y+2z=14$.

5. Find the distance between the parallel planes $2x-2y+z+1=0$ and $4x-4y+2z+3=0$.

6. Find the equation of the plane passing through the line of intersection of the planes $2x-y=0$ and $3z-y=0$, and perpendicular to the plane $4x+5y-3z=8$. (*A.I.S.S.C.E., 1985*)

7. Find the equation of the plane passing through the origin and the line of intersection of the plannes
$x+2y+3z+4=0$ and $x-y+z+3=0$ (*A.I.S.S.C.E., 1987*)

8. Find the equation of the plane passing through the intersection of the planes $2x+y+2z=9$, $4x-5y-4z=1$, and the point (3, 2, 1).

9. Find the equation of the plane passing through the line of intersection of the planes $x+y+z=6$ and $2x+3y+4z+5=0$ and perpendicular to the plane $4x+5y-3z=8$.

10. Find the equations of the plane passing through the line of intersection of the planes $x+3y+6=0$, $3x-y-4z=0$ and whose perpendicular distance from the origin is unity.

11. Find the equation of the plane through the intersection of the planes $x+y+z=1$ and $2x+3y-z+4=0$ which is parallel to the x-axis.

12. Show that the line

$$\frac{x+10}{1}=\frac{y-8}{-2}=\frac{z}{1}$$

lies in the plane $x+2y+3z-6=0$.

13. Show that the line

$$\frac{x-2}{3}=-(y+2)=\frac{1}{4}(z-3)$$

lies in the plane

$$2x+2y-z+3=0.$$

14. Show that the equation of the plane parallel to the join of $(3, 2, -5)$ and $(0, -4, -11)$ and passing through the points $(-2, 1, -3)$ and $(4, 3, 3)$ is $4x+3y-5z=10$.

15. Find the equation of the plane through the points $(1, 0, -1)$, $(3, 2, 2)$ and parallel to the line

$$\frac{x-1}{1}=\frac{y-1}{-2}=\frac{z-2}{3}$$

16. Find the equation of the plane containing the line

$$2x-5y+2z=6,\ 2x+3y-z=5$$

and parallel to the line $x=-y/6=z/7$.

17. Find the equation of the plane through the line of intersection of the planes $\mathbf{r}\cdot\mathbf{n}_1=1$, $\mathbf{r}\cdot\mathbf{n}_2=1$ and perpendicular to the plane $\mathbf{r}\cdot\mathbf{n}_3=1$.

18. Find the plane which passes through the point $\mathbf{a}$ and is perpendicular to the two planes $\mathbf{r}\cdot\mathbf{n}_1=q_1$, $\mathbf{r}\cdot\mathbf{n}_2=q_2$.

19. Find the equation of the plane which passes through the points $\mathbf{a}$ and $\mathbf{b}$, and is perpendicular to the plane $\mathbf{r}\cdot\mathbf{n}=q$.

9·6. THE SPHERE

Sphere has attracted the mathematicians since ancient times because the earth is nearly spherical. The heavenly bodies—the sun and the moon, are also spherical. Because of its symmetrical shape, the sphere has a lot of applications in our lives. In the present section we shall study the equation of a sphere.

A sphere is the locus of a point which moves so that its distance from a fixed point is constant. The fixed point is called the *centre* of the sphere and the fixed distance is called the *radius* of the sphere.

9·6 1. The Standard Form of the Equation of a Sphere.

Let the centre of a sphere be at the origin of reference, and let its radius be a. Let p be any point on the sphere. If $\mathbf{r}$ be the position vector of P, then

$$OP=a,$$

$$\Leftrightarrow \quad |\mathbf{r}|=a.$$

Therefore *the vectorial equation of a sphere whose centre is at the origin and radius is a is*

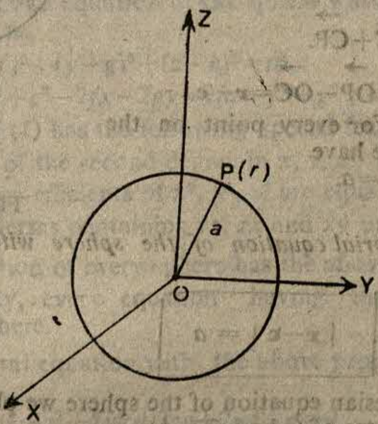


Fig. 9·14.

$$\boxed{|\mathbf{r}|=a} \qquad \text{...(A}$$

To obtain the equation of the sphere in cartesian form, we observe that if P be the point (x, y, z), then

$OP=\sqrt{((x-0)^2+(y-0)^2+(z-0)^2)\}}=\sqrt{(x^2+y^2+z^2)}$,

so that the required equation is

$$\sqrt{x^2+y^2+z^2}=a,$$

or $\qquad x^2+y^2+z^2=a^2.$

Therefore *the cartesian equation of a sphere with centre at the origin and radius a is*

$$x^2+y^2+z^2=a^2.$$

Example 25. *Find the cartesian equation of the sphere whose centre is at the origin and radius is 3 units.*

Solution. The equation of the sphere is

$\sqrt{\{(x-0)^2+(y-0)^2+(z-0)^2\}}=3,$

or $\quad x^2+y^2+z^2=9.$

9·6·2. Equation of the Sphere whose Centre and Radius are Given

Let a be the radius, and C be the centre of a sphere. Let the position vector of C be $\mathbf{c}$.

If $\mathbf{r}$ be the position vector of any point P on the sphere, then

$\vec{OC}=\mathbf{c}$, $\vec{OP}=\mathbf{r}$.

Now $\vec{OP}=\vec{OC}+\vec{CP}$.

so that $\vec{CP}=\vec{OP}-\vec{OC}=\mathbf{r}-\mathbf{c}$.

Since $CP=a$ for every point on the sphere, therefore we have

$$|\mathbf{r}-\mathbf{c}|=a.$$

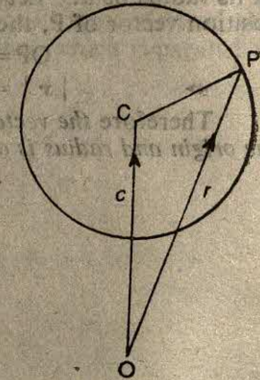


Fig. 9·15.

Thus *the vectorial equation of the sphere with centre C and radius a* is

$$\boxed{|\mathbf{r}-\mathbf{c}|=a} \qquad ...(A)$$

To obtain the cartesian equation of the sphere we observe that if the cartesian co-ordinates of C be (f, g, h),

then

$$\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k},$$
$$\mathbf{c}=f\mathbf{i}+g\mathbf{j}+h\mathbf{k}$$

so that $\mathbf{r}-\mathbf{c}=(x-f)\,\mathbf{i}+(y-g)\,\mathbf{j}+(z-h)\,\mathbf{k}$.

Substituting the above expression for $\mathbf{r}-\mathbf{c}$ in (A), we have

$$|(x-f)\,\mathbf{i}+(y-g)\,\mathbf{j}+(z-h)\,\mathbf{k}|=a,$$

or $$\boxed{(x-f)^2+(y-g)^2+(z-h)^2=a^2} \qquad ...(B)$$

as the desired equation of the sphere with centre (f, g, h) and radius a.

Remark. We could have obtained (B) directly with using (A). In fact, by the formula for the distance between two points,

$$CP=a$$
$$\Leftrightarrow \quad \sqrt{\{(x-f)^2+(y-g)^2+(z-h)^2\}}=a$$
$$\Leftrightarrow \quad (x-f)^2+(y-g)^2+(z-h)^2=a^2.$$

Example 26. *Find the vectorial equation of the sphere with centre $2i-3j+4k$ and radius 5 units.*

Solution. Here $\mathbf{c}=2i-3j+4k$, $a=5$.

Therefore the equation is

$$|\mathbf{r}-(2\mathbf{i}-3\mathbf{j}+4\mathbf{k})|=5. \qquad \text{...(1)}$$

The cartesian form of (1) is

$$\sqrt{\{(x-2)^2+(y+3)^2+(z-4)^2\}}=5,$$

or $$(x-2)^2+(y+3)^2+(z-4)^2=25,$$

or $$x^2+y^2+z^2-4x+6y-8z+4=0. \qquad \text{...(2)}$$

9·6·3. The General Equation of a Sphere.

The cartesian equation of the sphere whose centre is (f, g, h) and radius is a is

$$(x-f)^2+(y-g)^2+(z-h)^2=a^2,$$

or $$x^2+y^2+z^2-2fx-2gy-2hz+f^2+g^2+h^2-a^2=0. \qquad \text{...(1)}$$

Equation (1) has the following special features :

(*i*) It is of the second degree in x, y, z.

(*ii*) The co-efficients of x^2, y^2, z^2 are equal.

(*iii*) The terms containing yz, zx and xy are absent.

The equation of every sphere has the above properties.

Conversely, every equation having the above properties represents a sphere.

The general equation with the above properties can be written as

$$x^2+y^2+z^2+2ux+2vy+2wz+d=0. \qquad \text{...(2)}$$

Equation (2) may be re-written as

$$(x^2+2ux+u^2)+(y^2+2vy+v^2)+(z^2+2wz+w^2) \\ =u^2+v^2+w^2-d$$

or $$(x+u)^2+(y+v)^2+(z+w)^2=u^2+v^2+w^2-d,$$

showing that the distance of the point (x, y, z) from the fixed point $(-u, -v, -w)$ is $\sqrt{(u^2+v^2+w^2-d)}$. The locus of (x, y, z) is, therefore, a sphere whose centre is the point $(-u, -v, -w)$ and whose radius is $\sqrt{(u^2+v^2+w^2-d)}$.

Thus, *the equation*

$$x^2+y^2+z^2+2ux+2vy+2wz+d=0$$

rapresents a sphere whose centre is the point $(-u, -v, -w)$ *and radius is* $\sqrt{(u^2+v^2+w^2-d)}$.

Rule. *If the equation of a sphere be*

$$x^2+y^2+z^2+2ux+2vy+2wz+d=0,$$

the co-ordinates of the centre

$$=(-\tfrac{1}{2}\text{ co-eff. of } x, -\tfrac{1}{2}\text{ co-eff. of } y, -\tfrac{1}{2}\text{ co-eff. of } z),$$

$$\text{radius}=\sqrt{\{(\tfrac{1}{2}\text{ co-eff. of } x)^2+(\tfrac{1}{2}\text{ co-eff of } y)^2+(\tfrac{1}{2}\text{ co-eff. of } z)^2 \\ -\text{constant term}\}}.$$

Remarks 1. The general equation of a sphere in vector form is

$$\mathbf{r}^2-2\mathbf{r}.\mathbf{c}+k=0,$$

where $\mathbf{c}$ is a given vector, and k is a given scalar.

2. For equation (2) to represent a sphere we must have $u^2+v^2+w^2-d>0$.

Example 27. *Find the radius and co-ordinates of the centre of the sphere whose equation is*

$$3x^2+3y^2+3z^2-6x-12y+6z+2=0 \qquad \textit{(A.I.S.S.C.E., 1984)}$$

Solution. Dividing the given equation throughout by 3, we can write it as

$$x^2+y^2+z^2-2x-4y+2z+\frac{2}{3}=0. \qquad \ldots(1)$$

Comparing (1) with

$$x^2+y^2+z^2+2ux+2vy+2wz+d=0,$$

we have $2u=-2, 2v=-4, 2w=2, d=\dfrac{2}{3}$.

or $\qquad u=-1, \ v=-2, \ w=1, \ d=\dfrac{2}{3}$.

Hence the centre of the sphere is the point $(-u, -v, -w)$, *i.e.*, $(1, 2, -1)$, and the radius of the sphere

$$=\sqrt{(u^2+v^2+w^2-d)},$$

$$=\sqrt{\left\{(-1)^2+(-2)^2+1^2-\frac{2}{3}\right\}},$$

$$=\sqrt{\frac{16}{3}}=\frac{4\sqrt{3}}{3}.$$

Aliter. The given equation can be written as

$$x^2+y^2+z^2-2x-4y+2z+\frac{2}{3}=0,$$

or $$(x^2-2x+1)+(y^2-4y+4)+(z^2+2z+1)=1+4+1-\frac{2}{3},$$

or $$(x-1)^2+(y-2)^2+(z+1)^2=\frac{16}{3}$$

Therefore the centre is $(1, 2, -1)$ and radius is $\sqrt{\dfrac{16}{3}}$ *i.e.*, $\dfrac{4\sqrt{3}}{3}$.

Example 28. *Find the equation of the sphere concentric with*

$$x^2+y^2+z^2-2x-4y-6z-11=0$$

but of double the radius. (D.B.S.S.C.E., 1986)

Solution. The equation

$$x^2+y^2+z^2-2x-4y-6z-11=0$$

can be written as

$$(x-1)^2+(y-2)^2+(z-3)^2=25,$$

so that the centre of the given sphere is the point (1, 2, 3) and its radius is 5 units. We are required to find the equation of the sphere whose centre is (1, 2, 3) and whose radius is 10 (=twice of 5) units. The required equation is

$$(x-1)^2+(y-2)^2+(z-3)^2=10^2,$$

or $$x^2+y^2+z^2-2x-4y-6z-86=0.$$

9·6·4. The Equation of the Sphere when the Position Vectors (or co-ordinates) of the Extremities of a Diameter are Given

Let **a** and **b** be the position vectors of the extremities A and B of a diameter of a sphere and let P be any point on the sphere.

If the position vector of P with respect to O as the origin of reference be **r**, then

$$\vec{OA}=\mathbf{a},\ \vec{OB}=\mathbf{b},\ \vec{OP}=\mathbf{r},$$

so that

$$\vec{AP}=\vec{OP}-\vec{OA}=\mathbf{r}-\mathbf{a},$$

$$\vec{BP}=\vec{OP}-\vec{OB}=\mathbf{r}-\mathbf{b}.$$

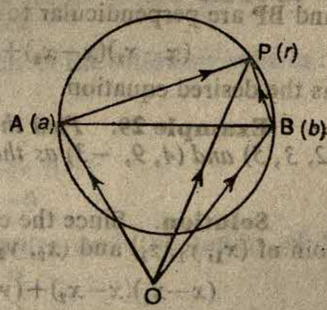


Fig. 9·16.

Since A and B are the extremities of a diameter, and P is a point on the sphere, therefore

$\angle APB$ is a right angle, so that

$$\vec{AP}\cdot\vec{BP}=0,$$

or $$(\mathbf{r}-\mathbf{a})\cdot(\mathbf{r}-\mathbf{b})=0,$$

is the required equation.

Thus, *the vectorial equation of the sphere with* **a** *and* **b** *as the position vectors of the extremities of a diameter is*

$$\boxed{(\mathbf{r}-\mathbf{a})\cdot(\mathbf{r}-\mathbf{b})=0} \qquad ...(A)$$

To obtain the cartesian equation of the sphere having (x_1, y_1, z_1) and (x_2, y_2, z_2) as the extremities of a diameter we have to substitute

$$\mathbf{a}=x_1\mathbf{i}+y_1\mathbf{j}+z_1\mathbf{k},$$
$$\mathbf{b}=x_2\mathbf{i}+y_2\mathbf{j}+z_2\mathbf{k},$$
$$\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k},$$

in (A).

Now $\quad \mathbf{r}-\mathbf{a}=(x-x_1)\,\mathbf{i}+(y-y_1)\,\mathbf{j}+(z-z_1)\,\mathbf{k},$

$\quad\quad \mathbf{r}-\mathbf{b}=(x-x_2)\,\mathbf{i}+(y-y_2)\,\mathbf{j}+(z-z_2)\,\mathbf{k},$

so that

$$(\mathbf{r}-\mathbf{a})\,.\,(\mathbf{r}-\mathbf{b})=0.$$
$$\Leftrightarrow \quad (x-x_1)(x-x_2)+(y-y_1)(y-y_2)+(z-z_1)(z-z_2)=0.$$

Thus *the equation of the sphere with* (x_1, y_1, z_1) *and* (x_2, y_2, z_2) *as the extremities of a diameter is given by*

$$\boxed{(x-x_1)(x-x_2)+(y-y_1)(y-y_2)+(z-z_1)(z-z_2)=0.} \quad \text{...(B)}$$

Remark. We could have obtained the above equation directly by observing that the direction ratios of AP are $(x-x_1, y-y_1, z-z_1)$ and those of BP are $(x-x_2, y-y_2, z-z_2)$. Since AP and BP are perpendicular to each other, we have

$$(x-x_1)(x-x_2)+(y-y_1)(y-y_2)+(z-z_1)(z-z_2)=0$$

as the desired equation.

Example 29. *Find the equation of the sphere on the join of (2, 3, 5) and (4, 9, −3) as the extremities of a diameter.*

(A.I.S.S.C.E., 1984)

Solution. Since the equation of the sphere described on the join of (x_1, y_1, z_1) and (x_2, y_2, z_2) as the extremities of a diameter is

$$(x-x_1)(x-x_2)+(y-y_1)(y-y_2)+(z-z_1)(z-z_2)=0,$$

therefore the equation of the sphere described on the join of (2, 3, 5) and (4, 9, −3) as the extremities of a diemeter is

$$(x-2)(x-4)+(y-3)(y-9)+(z-5)(z+3)=0,$$

or $\quad (x^2-6x+8)+(y^2-12x+27)+(z^2-2z-15)=0,$

or $\quad x^2+y^2+z^2-6x-12y-2z+20=0.$

Aliter. The centre of the sphere is the mid-point of the join of (2, 3, 5) and (4, 9, −3). Therefore the centre is the point

$$\left(\frac{2+4}{2}, \frac{3+9}{2}, \frac{5+(-3)}{2}\right), \text{ i.e., } (3, 6, 1).$$

Diameter of the sphere $=\sqrt{\{(4-2)^2+(9-3)^2+(-3-5)^2\}}$,

$$=2\sqrt{26}.$$

$$=\sqrt{26}.$$

$\therefore$ Radius

Hence the equation of the sphere is

$$(x-3)^2+(y-6)^2+(z-1,^2=(\sqrt{26})^2,$$

or $$x^2+y^2+z^2-6x-22y-2z+20=0.$$

EXERCISE 9 (*h*)

1. Find the cartesian equation of the sphere whose
   (*i*) centre is (1, 0, 0) and radius is 2 units ;
   (*ii*) centre is (0, 1, −1) and radius is 3 units ;
   (*iii*) centre is (2, −1, 3) and radius is 5 units.
2. Find the vectorial equation of the sphere with
   (*i*) centre $3\mathbf{j}$ and radius 2 units ;
   (*ii*) centre $\mathbf{i}-\mathbf{j}$ and radius 4 units ;
   (*iii*) centre $\mathbf{i}+\mathbf{j}-\mathbf{k}$ and radius 8 units.
3. Find the cartesian equation of the sphere whose centre has position vector $3\mathbf{i}+2\mathbf{j}-\mathbf{k}$ and whose radius is 6 units.
4. Find the vectorial equation of the sphere whose centre is the point (1, 2, 3) and radius is 4 units.
5. Find the centre and radius of each of the following spheres :
   (*i*) $x^2+y^2+z^2=25.$
   (*ii*) $x^2+y^2+z^2-2x+4y=0.$
   (*iii*) $x^2+y^2+z^2+6x-8y+10z-14=0.$
   (*iv*) $x^2+y^2+z^2+4x+6y-8z-7=0.$
6. Find the position vector of the centre and the radius of the sphere whose cartesian equation is

$$x^2+y^2+z^2-6z-9=0.$$

7. Find the co-ordinates of the centre and radius of the sphere whose vector equation is

$$\mathbf{r}^2-\mathbf{r}\,.\,(8\mathbf{i}-6\mathbf{j}+10\mathbf{k})-50=0.$$

8. Find the equation of the sphere having (1, 2, 3) and (−1, −2, −3) as the extremities of a diameter.
9. Find the vectorial equation of the sphere having $\mathbf{i}+\mathbf{j}$ and $\mathbf{j}+\mathbf{k}$ as the extremities of its diameter. Find its centre and radius.
10. Find the cartesian equation of the sphere having $2\mathbf{i}+\mathbf{j}$ and $3\mathbf{j}-8\mathbf{k}$ as the position vectors of the extremities of a diameter. Find its centre and radius.

11. If A $(-1, 4, -3)$ is one end of a diameter AB of the sphere

$$x^2+y^2+z^2-3x-2y+2z-15=0.$$

then find the co-ordinates of B, the other end.

(*A.I.S.S.C.E.*, *1988*)

12. Find the centre and radius of the sphere

$$5x^2+5y^2+5z^2+10x-6y+8z+5=0.$$

(*D.B.S.S.C.E.*, *1988*)

13. Find the equation of a sphere which passes through the points $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ and whose centre lies on the plane $3x-y+z=2$. (*A.I.S.S.C.E.*, *1985*)

14. Find the centre and radius of the sphere

$$(x-1)(x+2)+(y-2)(y+4)+(z-3)(z+6)=0$$

(*A.I.S.S.C.E.*, *1985*)

15. Obtain the equation of the sphere passing through the points $(1, 2, 3)$, $(0, 3, 3)$, $(1, 3, 2)$ and having its centre on the plane $x+4y+z=0$. (*A.I.S.S.C.E.*, *1986*)

16. Find the equation of the sphere whose centre is the point $(2, 3, 1)$ and which touches the plane $x+y+z=0$.

(*A.I.S.S.C.E.*, *1986*)

17. Find the equation of the sphere concentric with $x^2+y^2+z^2-2x-4y-6z-11=0$ but of half the radius.

18. Find the radius and centre of the sphere

$$4x^2+4y^2+4z^2-4x+6y-8z-2=0.$$

(*D.B.S.S.C.E.*, *1987*)

TEST YOUR UNDERSTANDING IX

In each of the following problems, four alternatives are given out of which one is correct. Put a tick-mark (✓) against the correct alternative.

1. The distance between the points $(1, -2, 3)$, $(-1, 2, -3)$ is

(*a*) $\sqrt{14}$ units (*b*) $\sqrt{12}$ units

(*c*) $\sqrt{14}$ units (*d*) $\sqrt{28}$ units.

2. The co-ordinates of the mid-point of the line segment joining P $(2, -5, a)$ and $(4, 9, 3)$ are $(3, 2, -7)$. The value of a is

(*a*) -4 (*b*) 17

(*c*) -10 (*d*) -17.

3. The direction ratios of two lines are $(1, -3, 2)$ and $(k, 2, 4)$. If the lines are perpendicular to each other, the value of k is

(a) 2 (b) -2
(c) -6 (d) 6.

4. The direction ratios of two lines are $(-1, 8, 6)$ and $(2, t, -12)$. If the lines are parallel to each other, the value of t is
(a) 16 (b) -16
(c) -4 (d) 8.

5. The angle between the planes $x+y=0$ and $y+z=0$ is
(a) $\pi/3$ (b) $\pi/4$
(c) $\pi/6$ (d) $\pi/2$.

6. The planes $x+2y-2z=1$ and $2x+ky+2z=3$ are perpendicular to each other. The value of k is
(a) -1 (b) 1
(c) 2 (d) -2.

7. The plane $ax+2y-3z=4$ is parallel to the line $\frac{x}{1}=\frac{y}{2}=\frac{z}{-3}$. The value of a is
(a) 13 (b) -1
(c) 1 (d) -13.

8. The plane $5x-3y+4z=1$ is perpendicular to the line $\frac{x}{k}=\frac{y}{3}=\frac{z}{-4}$. The value of k is
(a) 5 (b) -4
(c) -5 (d) 3.

9. The co-ordinates of the centre of a sphere are $(-1, 2, -3)$. If one extremity of a diameter is $(5, -8, 3)$ and the other extremity is $(-7, 12, k)$, the value of k is
(a) 3 (b) -3
(c) -9 (d) -6.

10. The diameter of the sphere $x^2+y^2+z^2-6x-8y+10z+k=0$ is 10 units. The value of k is
(a) 25 (b) -50
(c) 0 (d) 100.

REVIEW EXERCISE IX

1. Find the direction cosines of the line joining the points P $(4, 3, -5)$ and Q $(-2, 1, -8)$. (*D.B.S.S.C.E., 1984*)

2. Find the projection of the line segment joining the points $(-1, 0, 3)$ and $(2, 5, 1)$ on the line whose direction ratios are $(6, 3, 2)$. (*D.B.S.S.C.E.*, *1985*)

3. Show that the two lines

$$\frac{x-1}{2}=\frac{y-2}{3}=\frac{z-3}{4}$$

and

$$\frac{x-4}{5}=\frac{y-1}{2}=z$$

intersect. Find also the point of intersection. (*A.I.S.S.C.E.*, *1984*)

4. Show that the line

$$\frac{x-3}{3}=\frac{y-2}{-4}=\frac{z+1}{1}$$

intersects the line

$$2x+2y+3z=0,\ 2x+4y+3z-60=0.$$

Find the co-ordinates of the point of intersection. (*A.I.S.S.C.E.*, *1988*)

5. Find the equation of the perpendicular drawn from the point $(2, 4, -1)$ to the line

$$\frac{x+5}{1}=\frac{y+3}{4}=\frac{z-6}{-9} \text{ and}$$

obtain the co-ordinates of the foot of the perpendicular. (*D.B.S.S.C.E.*, *1988*)

6. Find the co-ordinates of two points on the line given by the equations

$$\frac{x-1}{2}=\frac{y-2}{3}=\frac{z-3}{4}$$

at a distance $\frac{\sqrt{29}}{2}$ from the point $(1, 2, 3)$. (*A.I.S.S.C.E.*, *1986*)

7. The vertices of a triangle are A $(3, 2, 4)$, B $(6, 12, 5)$, C $(11, 6, 8)$. Find the angles of the triangle ABC.

8. A plane meets the co-ordinate axes at A, B, C such that the centroid of the triangle ABC is at the point (a, b, c). Show that the equation of the plane is

$$\frac{x}{a}+\frac{y}{b}+\frac{z}{c}=3.$$

9. Show that the equation of the plane passing through the line $u_1 \equiv a_1x+b_1y+c_1z+d_1=0$, $u_2 \equiv a_2x+b_2y+c_2z+d_2=0$ and parallel to the line

$$\frac{x}{l}=\frac{y}{m}=\frac{z}{n}$$

is

$u_1\,(a_2l+b_2m+c_2n)=u_2\,(a_1l+b_1m+c_1n).$

10. Find the perpendicular distance of the point $(2, -7, 15)$ from the line

$$\frac{x}{2}=\frac{y+2}{l}=\frac{z-1}{2}$$

11. Find the point in which the line

$$\frac{x-1}{2}=\frac{y-2}{-3}=\frac{z+3}{4}$$

meets the plane $2x+3y-z+1=0$.

12. Find the distance of the point $(1, -2, 3)$ from the plane $x-y+z=5$ measured parallel to the line

$$\frac{x}{2}=\frac{y}{3}=\frac{z}{-6}$$

13. Find the distance between the parallel planes

$$3x+4y-12z=3,\ 6x+8y-24z+9=0.$$

14. Find the distance of the point $(2, 3, -5)$ from the plane

$$x+2y-2z=15.$$

15. Find the equation of the plane passing through the intersection of the planes $7x+2y+3z-4=0$ and $4x+3y+2z+1=0$ and passing through the origin.

16. Show that the four points $(0, -1, 0)$, $(2, 1, -1)$, $(1, 1, 1)$ and $(3, 3, 0)$ are coplanar.

17. Find the equation of the plane passing through the points $(2, -3, 1)$ and $(-1, 1, -7)$ and perpendicular to the plane $x-2y+5z+1=0$.

18. Find the equation of the plane passing through the points $(1, -1, -1)$ and perpendicular to each of the planes $x-2y-8z=0$ and $2x+5y-z=0$.

19 Find the equation of the sphere which has the line segment joining the points $(3, 1, -5)$ and $(1, -3, 1)$ as the extremities of a diameter. Find its centre and radius.

20. Find the equation of the sphere having its centre on the line of intersection of the planes $2x-3y=0$, $5y+2z=0$, and passing through the points $(0, -2, -4)$ and $(2, -1, -1)$.

SUMMARY

1. (a) Distance between the points with position vectors $\mathbf{r}_1, \mathbf{r}_2 = |\mathbf{r}_1 - \mathbf{r}_2|$.

   (b) Distance between the points $A(x_1, y_1, z_1)$, $B(x_2, y_2, z_2)$
   $= \sqrt{\{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2\}}$.

2. (a) The position vector of the point which divides the join of $P(\mathbf{r}_1)$ and $Q(\mathbf{r}_2)$ in the rario $n : m$ is

$$\frac{m\mathbf{r}_1+n\mathbf{r}_2}{m+n}.$$

   (b) The co-ordinates of the point dividing the join of (x_1, y_1, z_1) and (x_2, y_2, z_2) in the ratio $n : m$ are

$$\left(\frac{mx_1+nx_2}{m+n}, \frac{my_1+ny_2}{m+n}, \frac{mz_1+nz_2}{m+n}\right).$$

3. (a) The acute angle θ between the lines parallel to the vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ is given by

$$\cos\theta = \left(\frac{|\mathbf{d}_1 \cdot \mathbf{d}_2|}{|\mathbf{d}_1|\ |\mathbf{d}_2|}\right).$$

   (b) The acute angle θ between the lines with direction cosines (l_1, m_1, n_1) and (l_2, m_2, n_2) is given by

$$\cos\theta = |l_1l_2+m_1m_2+n_1n_2|$$

   (b) The acute angle θ between the lines with direction ratios (a_1, b_1, c_1), (a_2, b_2, c_2) is given by

$$\cos\theta = \frac{|a_1a_2+b_1b_2+c_1c_2|}{\sqrt{(a_1^2+b_1^2+c_1^2)}\sqrt{(a_2^2+b_2^2+c_2^2)}}.$$

4. (a) The lines $\mathbf{r}=\mathbf{r}_1+s\mathbf{d}_1$, $\mathbf{r}=\mathbf{r}_2+t\mathbf{d}_2$ are parallel provided $\mathbf{d}_1 \times \mathbf{d}_2 = \mathbf{0}$.

   (b) The lines with direction ratios (l_1, m_1, n_2) and (l_2, m_2, n_2) are parallel provided

$$\frac{l_1}{l_2}=\frac{m_1}{m_2}=\frac{n_1}{n_2}.$$

5. (a) The lines $\mathbf{s}=\mathbf{r}_1+s\mathbf{d}_1$ and $\mathbf{r}=\mathbf{r}_2+t\mathbf{d}_2$ are perpendicular to each other provided $\mathbf{d}_1 \cdot \mathbf{d}_2=0$.

   (b) The lines with direction ratios (l_1, m_1, n_1), (l_2, m_2, n_2) are perpendicular to each other provided

$$l_1l_2+m_1m_2+n_1n_2=0.$$

6. (a) The parametric vectorial equation of a line through $\mathbf{r}_1$ parallel to the vector $\mathbf{d}$ is $\mathbf{r}=\mathbf{r}_1+t\mathbf{d}$, where t is a parameter.

   (b) The parametric vectorial equation of a line through the points $\mathbf{r}_1$, $\mathbf{r}_2$ is $\mathbf{r}=(1-t)\,\mathbf{r}_1+t\mathbf{r}_2$, where t is a parameter.

7. (a) The cartesian equation of the line with direction ratios (l, m, n), and passing through the point (x_1, y_1, z_1) is

$$\frac{x-x_1}{l}=\frac{y-y_1}{m}=\frac{z-z_1}{n}.$$

   (b) The cartesian equation of the line passing through the points $(x_1\ y_1, z_1)$ and (x_2, y_2, z_2) is

$$\frac{x-x_1}{x_2-x_1}=\frac{y-y_1}{y_2-y_1}=\frac{z-z_1}{z_2-z_1}.$$

8. (*a*) The shortest distance between the lines

$$\mathbf{r}=\mathbf{r}_1+s\mathbf{d}_1,\ \mathbf{r}=\mathbf{r}_2+t\mathbf{d}_2 \text{ is}$$
$$[\mathbf{r}_2-\mathbf{r}_1\ \mathbf{d}_1\ \mathbf{d}_2]/\ |\ \mathbf{d}_1\times\mathbf{d}_2\ |\ .$$

(*b*) The shortest distance between the lines

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1},$$
$$\frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}=\frac{z-z_2}{n_2},$$

is
$$\begin{vmatrix} x_2-x_1 & y_2-y_1 & z_2-z_1 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix} \div\sqrt{\Sigma(m_1n_2-m_2n_1)^2}.$$

9. (*a*) The vectorial equation of the plane normal to the unit vector $\mathbf{n}$ and at a distance p from the origin is $\mathbf{r}\ .\ \mathbf{n}=p$.

(*b*) The cartesian equation of the plane normal to a line with direction cosines (l, m, n) and at a distance p from the origin is $lx+my+nz=p$.

10. (*a*) The vectorial equation of the plane through the point $\mathbf{r}_1$ and normal to the vector $\mathbf{n}$ is $(\mathbf{r}-\mathbf{a}_1)\ .\ \mathbf{n}=0$.

(*b*) The cartesian equation of the plane through the point (x_1, y_1, z_1), and with (a, b, c) as the direction ratios of the normal to it, is

$$a(x-x_1)+b(y-y_1)+c(z-z_1)=0.$$

11. (*a*) The vectorial equation of the plane through the points $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ is

$$[\mathbf{r}-\mathbf{r}_1\ \ \mathbf{r}_2-\mathbf{r}_1\ \ \mathbf{r}_3-\mathbf{r}_1]=0.$$

(*b*) The cartesian equation of the plane through the points (x_1, y_1, z_1), (x_2, y_2, z_2) and (x_3, y_3, z_3) is

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \end{vmatrix}=0.$$

12. (*a*) The length of the perpendicular to the plane $\mathbf{r}\ .\ \mathbf{n}=q$ from the point $\mathbf{r}_1$ is $|\ q-\mathbf{r}_1\ .\ \mathbf{n}\ |\ /\ |\ \mathbf{n}\ |$.

(*b*) The length of the perpendicular to the plane $ax+by+cz+d=0$ from the point (x_1, y_1, z_1) is

$$\frac{|\ ax_1+by_1+cz_1+d\ |}{\sqrt{a^2+b^2+c^2}}.$$

13. (*a*) The angle θ between the planes $\mathbf{r}\ .\ \mathbf{n}_1=q_1$ and $\mathbf{r}\ .\ \mathbf{n}_2=q_2$ is given by

$$\cos\theta=\frac{|\ \mathbf{n}_1\ .\ \mathbf{n}_2\ |}{|\ \mathbf{n}_1\ |\ \ |\ \mathbf{n}_2\ |}.$$

(*b*) The acute angle θ between the planes
$a_1x+b_1y+c_1z+d_1=0$, $a_2x+b_2y+c_2z+d_2=0$ is given by

$$\cos\theta=\frac{|\ a_1a_2+b_1b_2+c_1c_2\ |}{\sqrt{(a_1^2+b_1^2+c_1^2)}\ \sqrt{(a_2^2+b_2^2+c_2^2)}}.$$

14. (*a*) The planes $\mathbf{r}\ .\ \mathbf{n}_1=q_1$ and $\mathbf{r}\ .\ \mathbf{n}_2=q_2$ are parallel provided $\mathbf{n}_1\times\mathbf{n}_2=0$, and perpendicular provided $\mathbf{n}_1\ .\ \mathbf{n}_2=0$.

(*b*) The planes $a_1x+b_1y+c_1z+d_1=0$, $a_2x+b_2y+c_2z+d_2=0$ are parallel provided

$\frac{a_1}{a_2}=\frac{b_1}{b_2}=\frac{c_1}{c_2}$, and perpendicular provided

$$a_1a_2+b_1b_2+c_1c_2=0.$$

15. (*a*) The angle θ between $\mathbf{r}=\mathbf{r}_1+t\mathbf{d}$ and the plane $\mathbf{r}\cdot\mathbf{n}=p$ is given by

$$\sin\theta=\frac{|\mathbf{n}\times\mathbf{d}|}{|\mathbf{n}|\cdot|\mathbf{d}|}.$$

(*b*) The angle θ between the line

$$\frac{x-x_1}{l}=\frac{y-y_1}{m}=\frac{z-z_1}{n}.$$

and the plane $ax+by+cz+d=0$ is given by

$$\sin\theta=\frac{|al+bm+cn|}{\sqrt{(a^2+b^2+c^2)}\sqrt{(l^2+m^2+n^2)}}.$$

16. The line $\frac{x-x_1}{l}=\frac{y-y_1}{m}=\frac{z-z_1}{n}$ is parallel to the plane $ax+by+cz+d=0$ provided $al+bm+cn=0$. It lies in the plane if in addition we also have $ax_1+by_1+cz_1+d=0$.

17. (*a*) The vectorial equation of the sphere with centre at the origin and radius a is $|\mathbf{r}|=a$.

(*b*) The cartesian equation of the sphere with centre at the origin and radius a is $x^2+y^2+z^2=a^2$.

18. (*a*) The vectorial equation of the sphere with centre $\mathbf{c}$ and radius a is $|\mathbf{r}-\mathbf{c}|=a$.

(*b*) The cartesian equation of the sphere with centre (f, g, h) and radius a is

$$(x-f)^2+(y-g)^2+(z-h)^2=a^2.$$

10. (*a*) The centre of speere $\mathbf{r}^2-2\mathbf{r}\cdot\mathbf{c}+q=0$ is the point $\mathbf{c}$ and radius is $\sqrt{(q-\mathbf{c}^2)}$.

(*b*) The centre of the sphere

$$x^2+y^2+z^2+2ux+2vy+2wz+d=0$$

is $(-u, -v, -w)$ and its radius is $\sqrt{(u^2+v^2+w^2-d)}$.

20. (*a*) The vectorial equation of the sphere with $\mathbf{a}$ and $\mathbf{b}$ as the extremities of a diameter is $(\mathbf{r}-\mathbf{a})\cdot(\mathbf{r}-\mathbf{b})=0$.

(*b*) The cartesian equation of the sphere with (x_1, y_1, z_1) and (x_2, y_2, z_2) as the extremities of a diameter is

$$(x-x_1)(x-x_2)+(y-y_1)(y-y_2)+(z-z_1)(z-z_2)=0.$$

HISTORICAL NOTE

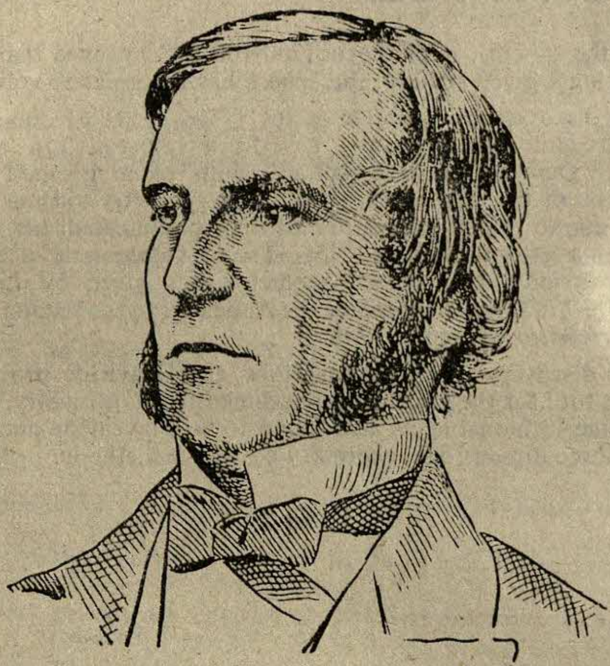
Geometry has fascinated the mathematicians for more than two thousand years. Euclid (365 B.C.) devoted the 12th book of his *Elements* to the study of the pyramid, cone and cylinder. Pythagoras (580-500 B.C.) called the sphere, the most beautiful of all solids. Archimedes (287 B.C.—212 B.C.) in his treatise *Sphere and the cylinder* proved among others the theorem that the surface of a sphere is four times the area of a great circle (*i.e.*, a circle obtained as a section of the sphere by a plane passing through the

centre of the sphere). Zenodorus (200-100 B.C.) proved that, of all solids having a given surface, the sphere has the greatest volume.

The discovery by Rene Descartes (1596-1650) of the analytic method for studying geometry gave fresh impetus to the study of geometry. Descartes dealt essentially with plane geometry. The introduction of the methods of co-ordinate geometry to three dimensions is due to John Bernoulli, who communicated his ideas to Leibnitz in a letter written in 1715. The first systematic account of three dimensional co-ordinate geometry was given by Leonhard Euler (1706-1783) in Chapter 5 of the Appendix to the second volume of his *Introductio* written in 1748.

The discovery of vectors by Gibbs and Heaviside provided an important tool for the study of three dimensional geometry. It has now become fashionable, as also found to be convenient and useful, to study three dimensional geometry by vector methods.

□ □



GEORGE BOOLE (1815-1864)

In the class-conscious society of the 19th century England, a boy by the name of George Boole was born at Lincoln in the meanest class—if at all it was considered a class. Boole aspired to belong to the upper echelons of the society. At the early age of eight, Boole determined to teach himself Greek and other languages reasonably well. His second and third attempts towards a dominant living were, respectively, taking a commercial course and trying to become a clergyman. Neither served the end to which it was supposed to be a means.

What did get Boole started was the early instruction in mathematics which he had received from his father—a petty shopkeeper. Around 1835 Boole opened a school of his own. It was at this time that Boole realized the inadequacy of the existing text-books and started reading the masters like Abel, Galois, Laplace and Lagrange, all on his own.

His first great achievement was the discovery of invariants—a basis for general relativity. Boole's other most notable contribution to mathematics was his theory of mathematical or symbolic logic. Boole added logic to the domain of Algebra. He separated the symbols of the binary compositions from the quantities composed by the same and developed an algebra of these symbols themselves – an original and a daring step. He formed what is known as a Boolean algebra—an algebraic system used extensively in informatics.

His pamphlet 'The Mathematical Analysis of Logic' and his masterpiece 'An Investigation of the Laws of Thought on which are Based the Mathematical Theories of Logic and Probabilities' are all-time source books.

Boole died in 1864 of Pneumonia.

CHAPTER 10

Mathematical Logic

10·1. INTRODUCTION

Mathematics is often called the 'Queen of Sciences'. This assertion is true for more reasons than one. One of the most important reasons for the supremacy of mathematics over other sciences is its exactness. Like most other sciences, mathematics has a language of its own—a language that surpasses every other language in its precision and brevity. To one who is trained in mathematical language, it communicates just what is intended—neither less nor more. To be a worthwhile student of mathematics, one must be able to use mathematical language with ease, efficiency and confidence. Another reason for the supremacy of mathematics over other sciences is its method, which consists in deriving conclusions from known assertions with the use of valid arguments. The method of reasoning used in mathematics is called *mathematical logic.* It owes its origin to the British mathematician George Boole. In the computer age its importance has grown because of its applications to switching circuits.

10·2. STATEMENTS

A sentence of which it is meaningful to say whether it is true or false, is called a *statement* or a *proposition*. We shall use the words 'statement' and 'proposition' interchangeably throughout. A statement must be either true or false, but not both. A true statement is said to have a truth value T, and a false statement is said to have truth value F. Instead of using letters T and F to denote truth values, it is useful to use the symbols 1 and 0. A true statement being given the truth value 1, and a false statement being given the truth value 0. (Some authors use the opposite of it, but that really does not matter !). The use of 1 and 0 is preferred over that of T and F because of the relative ease in computation, specially in applications to switching circuits.

Illustrations

1. Each of the following sentences is a statement :

(*a*) 3 is less than 6.

(*b*) The earth revolves about the sun.

(*c*) Mercury is a star.

(*d*) New York is a city in India.

2. None of the following sentences is a statement :

(*a*) Thank God !

(*b*) This proposition is false.

(*c*) Good bye !

If a sentence is a statement (or a proposition), it should be declaratory. It cannot be imperative, interrogative or exclamatory.

Statements/propositions will be denoted by small letters p, q etc.

Example 1. *Which of the following sentences are statements ;*

(*a*) *Bright sunshine !*

(*b*) $2^{127}-1$ *is a prime.*

(*c*) *Sing a song please !*

(*d*) *Why are you so sad ?*

(*e*) *New Delhi is the capital of India.*

Solution : (*b*) and (*e*) are statements. (*a*) and (*c*) are exclamatory sentences and are not statements. (*d*) is an interrogative sentence, and is not a statement.

Example 2. *Write the truth value of each of the following statements :*

(*a*) *The moon revolves round the sun.*

(*b*) *The sum of the angles of a triangle is* $180°$.

(*c*) *Multiplication of matrices is not commutative.*

(*d*) *The domain of the function* f *defined by* $f(x)=1/x$ *is* **R**.

(*e*) e^x *is strictly increasing for all* x.

Solution. Statements (*a*) and (*b*) are false. The truth value of each of them is 0.

Statements (*b*), (*c*) and (*e*) are true. The truth value of each of them is 1.

Remark. A statement must be either true or false but not both. It may happen sometimes (but only sometimes that a sentence appears to be a statement, even though actually it is not).

In the sixth century B.C. Epimenides, the celebrated poet and prophet of Crete, is said to have made the now famous remark "All Cretans are liars". We shall consider this assertion of Epimenides in somewhat detail. Let us first of all write it in the form, "All statements made by Cretans are false". If we remember that

Epimenides, the author of this assertion was himself a Cretan, what can we say about the truth or falsity of this assertion. Let us argue as follows :

(1) All statements made by Cretans are false.

(2) Statement (1) was made by a Cretan.

(3) Therefore statement (1) must be false.

(4) Therefore all statements made by Cretans are not false.

Since (1) and (4) cannot be simultaneously true, therefore we have a contradiction. This means that assertion (1) is not a statement in the sense in which we have used that term.

EXERCISE 10 (*a*)

1. Which of the following sentences or phrases, represent statements :

   (*a*) Ice is cold.

   (*b*) Is the number 3 a prime ?

   (*c*) Give me the pencil.

   (*d*) The number 1 is a prime.

   (*e*) Beautiful red roses.

   (*f*) He is ugly.

   (*g*) Cry baby !

2. Write the truth value of each of the following statements :

   (*a*) Angle in a semi-circle is a right angle.

   (*b*) 17 is the square of an integer.

   (*c*) 2 is the only odd prime.

   (*d*) The product of four consecutive natural numbers is a perfect square.

10·3. USE OF VENN DIAGRAMS IN LOGIC

We are already familiar with the use of Venn diagrams in representing sets. Venn diagrams can often be used with advantage in representing classes satisfying a given proposition by means of a diagram. Consider the following statements about students studying in class XII in the CBSE system :

(1) Every student studying Physics also studies Mathematics.

(2) Every student studying Mathematics also studies Physics.

(3) All students studying Physics also study Mathematics but some of the students studying Mathematics do not study Physics.

(4) All students studying Mathematics also study Physics, but some students studying Physics do not study Mathematics.

(5) A student studies Physics if and only if he studies Mathematics.

(6) No student studies Physics as well as Mathematics.

(7) While some students study both Physics as well as Mathematics, there are some students who study Physics but not Mathematics, and there are some who study Mathematics but not Physics.

Let us denote by U the universal set, that is, the set of all students studying in class XII of the CBSE system. Let P represent the set of all students who study Physics, and let M represent the set of all students who study Mathematics.

Fig. 10.1 (*a*) represents the truth of the statement (1).

Fig. 10.1 (*b*) represents the truth of the statement (2).

Fig. 10.1 (*c*) represents the truth of the statement (3).

Fig. 10.1 (*d*) represents the truth of statement (4).

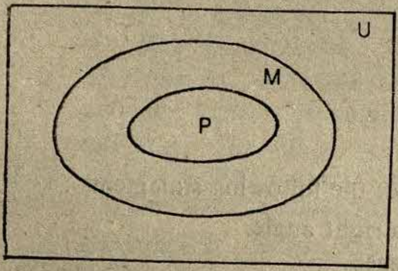


Fig. 10·1 (a)

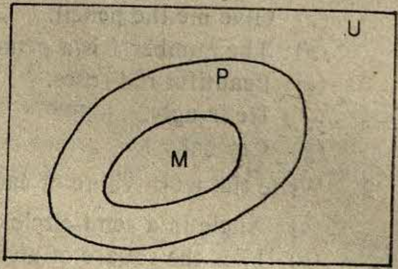


Fig. 10·1 (b)

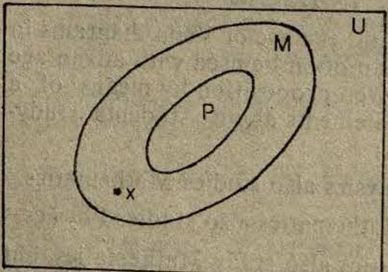


Fig. 10·1 (c)

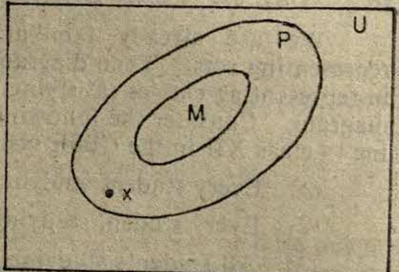


Fig. 10·1 (d)

The truth of the statement (5) is represented by Fig. 10.1 (e) and that of (6) by Fig. 10·1 (*f*).

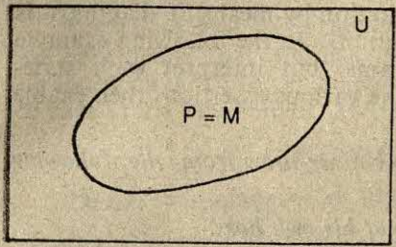


Fig. 10·1 (e)

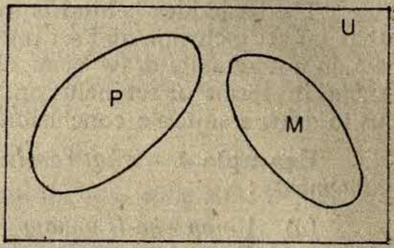


Fig. 10·1 (f)

The truth of statement (7) is represented by Fig. 10·1 (g).

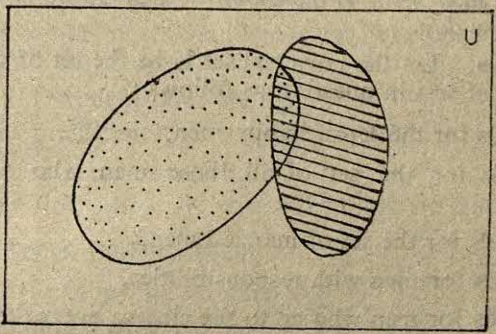


Fig. 10 1 (g)

Example 3. *Draw a Venn diagram to illustrate the statements :*

(1) *Socrates is a man.*

(2) *All men are mortal.*

What conclusion can you draw from the diagram ?

Solution. Let the universal set be the set U of all mortal beings, let M be the set of all men, and S the one-elementic set consisting of Socrates. Fig. 10·2 is the desired Venn diagram. From the diagram we find that $S \subset U$, *i.e.*, Socrates is mortal.

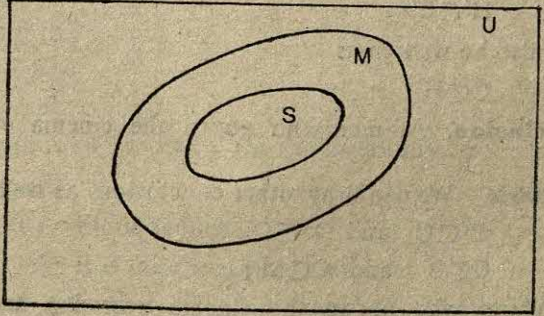


Fig. 10·2.

The basic idea behind representation by means of diagrams is that of set inclusion and set intersection. In the following example we do not actually draw Venn diagrams, but interpret each statement in terms of set inclusion. The various assertions then enable us to draw a suitable conclusion.

Example 4. *What conclusion can we draw from the following statements :*

(*a*) *A man who is unhappy is not his own boss.*

(*b*) *All married men have responsibilities.*

(*c*) *Every man is either his own boss or is married (or both).*

(*d*) *No man with responsibilities can go to see the cinema every day.*

Solution. Let the universal set U be the set of all men. Let us denote other sets in question as follows :

H stands for the set of happy men,

B stands for the set of all those men who are their own bosses,

M stands for the set of married men,

R stands for men with responsibilities,

C stands for men who go to the cinema every day.

Statement (a) : $H' \subset B'$

(b) : $M \subset R$

(c) : $B \cup M = U$

(d) : $R \subset C'$

Now (c) can be expressed alternatively

as (c^*) : $B' \subset M$

Combining (a), (c^*), (b), and (d), we have

$$H' \subset B' \subset M \subset R \subset C'$$

i.e., $H' \subset C'$

which can also be written as

$$C \subset H.$$

Conclusion. All men who go to the cinema everyday are happy.

Remark. We can draw other conclusions as well, *e.g.*,

$$B' \subset C' \text{ and } M \subset C',$$

i.e., $C \subset B$ and $C \subset M'$,

i.e., (1) Those who go to the cinema everyday are their own bosses.

(2) Those who go to the cinema everyday are not married. The conclusion that we had drawn first of all was the one which was based on *all* the given statements. The conclusions (1) and (2) are based only on some of the statements.

EXERCISE 10 (*b*)

Draw Venn diagrams to illustrate the following statements :

1. (*a*) All squares are rhombuses but some rhombuses are not squares.
   (*b*) All rectangles are parallelograms but some parallelograms are not rectangles.
   (*c*) All rhombuses are parallelograms but some parallelograms are not rhombuses.
2. (*a*) All IIT students are intelligent.
   (*b*) Some intelligent students study in IITs.
3. (*a*) All wise people are rich.
   (*b*) Some rich people are wise but others are not.

10·4. NEGATION OPERATION

Consider the following statements :

(*a*) Saurabh is intelligent.

(*b*) It is not true that Saurabh is intelligent.

Here the second statement has been formed from the first by prefixing it with the phrase 'It is not true that'. We say that the second statement is the negation of the first. If we denote the first statement by 'p', the second statement will be denoted by 'It is not true that p' or by 'It is false that p',' or more simply by 'not p'.

Definition 10·1. *Given any statement p, the negation of p is the statement 'It is false that p'.*

The negation of a statement p is denoted in any of the following ways :

(*i*) not 'p';

(*ii*) $\sim p$;

(*iii*) p'.

We shall prefer to use p' for the negation of p.

Illustrations.

1. Let p=blood is blue.

p'=It is false that blood is blue.

It is false that blood is blue can also be reworded as 'blood is not blue'.

2. Let $p = 2^{101}-1$ is a prime.

Then $p' = 2^{101}-1$ is not a prime.

3. Let p=Walking is good for health.

p'=It is false that walking is good for health.

p' can also be reworded as 'Walking is not good for health'. By the definition of the negation of a statement we find that if a statement p has the truth value 1, then p' has the truth value 0. If p has the truth value 0, then p' has the truth value 1. The truth values of p' corresponding to the possible truth values of p can be arranged in the form a table. Such a table is called a **truth table.** Truth tables will be found of use in writing truth values of statements in a systematic manner. In Table 10·1 we have given the truth table for negations.

TABLE 10.1

Truth table for negation

Row	p	p'
1	1	0
2	0	1

10·5. BASIC LOGIC CONNECTIVES AND COMPOUND STATEMENTS

Consider the following statements :

(*a*) The earth revolves round the sun.

(*b*) The number 1331 is the cube of an integer.

(*c*) $391581 \times 2^{216193}-1$ is the largest known prime.

(*d*) $2^{216091}-1$ is a prime.

(*e*) The sky is blue and the grass is green.

(*f*) Cars are costly *or* scooters are cheap.

(*g*) *If* 6789 is a perfect square, *then* the sum of the angles of a triangle is 180°.

(*h*) A quadrilateral is a rhombus *if and only if* the diagonals bisect each other at right angles.

The statements (*a*)–(*d*) are all simple statements. They cannot be split up into two or more statements. On the other hand, each of the statements (*e*)–(*h*) has been formed by connecting two statements with 'and', 'or', 'if...then', 'if and only if'. Such statements are called **compound statements**, and the words, word-pairs, phrases used for combining statements are called **connectives.** In this section we shall study the connectives 'and' and 'or' and the

compound statements formed by the operations 'disjunction' and 'conjunction'.

10·5·1. Disjunction

The connective '*or*' is applied to two statements p and q to form the compound statement 'p or q'. It called the *disjunction* of p and q. In symbols we write $p \vee q$, or $p+q$. We shall prefer to use $p+q$ for the disjunction of p and q.

Illustrations

1. Let p=It is raining.

   q=I am going for a walk.

   Then $p+q$ is the statement 'Either it is raining or I am going for a walk'.

2. Let p=He is handsome.

   q=He is intelligent.

   Then $p+q$ is the statement 'Either he is handsome or he is intelligent'.

In the English language the word 'or' is used in two different senses. Sometimes it means that exactly one of the two alternatives occurs, and sometimes it means that atleast one of the two alternatives occurs. In logic, the second of these two meanings is taken as standard. In other words, 'or' means the inclusive 'or' and not the exclusive 'or'. That is why the first row of the truth table (Table 10·2) show that when both p and q are true, then $p+q$ is true.

In Table 10·2 we have given the truth table for disjunction of p and q. Observe that there are four rows in this table since there are four possibilities for the pair of truth values of the statements p and q.

TABLE 10.2
Truth table for p + q

Row	p	q	p + q
1	1	1	1
2	1	0	1
3	0	1	1
4	0	0	0

The disjunction of p and q has the truth value 0 only when both p and q have truth value 0. In all other cases the truth value of $p+$q is 1.

10·5·2. Conjunction

The connective 'and' is applied to two statements p and q to form the compound statement 'p and q'. It is called the **conjunction** of p and q. In symbols, we use p & q, or $p \wedge q$, or pq to denote the conjunction of p and q. We shall prefer to use pq throughout.

Illustrations.

1. Let p=Steam is hot.
   q=The sky is blue.
   Then pq=Steam is hot and the sky is blue.
2. Let p=Mathematics is easy.
   q=I am intelligent.
   Then pq=Mathematics is easy and I am intelligent.

The following Table 10·3 gives the truth tables for the conjunction of p and q. Observe that there are four rows in this table since there are 4 possibilities for the pair of truth values of the statements p and q.

TABLE 10.3
Truth table for pq

Row	p	q	pq
1	1	1	1
2	1	0	0
3	0	1	0
4	0	0	0

The conjunction of p and q has the truth value 1 when both p and q have truth value 1. In all other cases the truth value of pq is 0.

Example 5. *Construct the truth table for $p+p'$ and pp'.*

Solution.

TABLE 10.4
Truth table for p + p' and pp'

Row	p	p'	p + p'	pp'
1	1	0	1	0
2	0	1	1	0

Remarks 1. From the above table we find that all the entries in the column headed by $p+p'$ are 1, *i.e.*, the statement $p+p'$ always has the truth value 1, whatever may be the truth value of p. A statement which is always true is called a **tautology.** We have seen above that $p+p'$ is a tautology. We shall use the symbol 1 to denote all the tautologies. Since it will be clear from the context as to whether 1 stands for a statement or for a truth value, no confusion is likely to arise.

2. From the above table we find that all the entries in the column headed by pp' are 0, *i.e.*, the statement pp' always has the truth value 0, whatever may be the truth value of p.

A statement which is always false is denoted by 0. Since it will be clear from the context as to whether 0 stands for a statement or for a truth value, no confusion is likely to arise.

In view of the above remarks we find that if p be *any* statement, then

$$p+p'=1,$$
$$pp'=0.$$

If we use the symbols $\sim$, $\vee$, and $\wedge$ instead of "′", '+' and '.' respectively, the above assertions can be expressed as follows :

$$\boxed{\begin{array}{c} p \vee (\sim p)=1 \\ p \wedge (\sim p)=0 \end{array}}$$

Theorem 10·1. *The operations of conjunction and disjunction on propositions are commutative.*

Proof. Construct the truth table for $p+q$, $q+p$, pq, qp.

TABLE 10.5
Truth table for p + q, q + p, pq, qp

Row	p	q	p + q	q + p	pq	qp
1	1	1	1	1	1	1
2	1	0	1	1	0	0
3	0	1	1	1	0	0
4	0	0	0	0	0	0

From the above table we find that

$$p+q=q+p$$
$$pq=qp$$

Remark. If we use the symbols $\vee$ and $\wedge$ instead of $+$ and juxtaposition respectively, the above statements can be expressed as follows :

$$\boxed{\begin{array}{c} p \vee q=q \vee p \\ p \wedge q=q \wedge p \end{array}}$$

Theorem 10·2. (*Existence of identity*).

There exist statements 1 and 0 respectively such that

$$p+0=0+p=p,$$
$$p1=1p=p,$$

for all statements p.

Proof. Follows immediately by constructing truth tables.

Theorem 10·3. *Each of the two operations (+) and (.) (where a.b is written as ab for the sake of simplicity) is distributive over the other. That is, if p, q, r be three statements, then*

(*a*) $p(q+r)=pq+pr$;

(*b*) $p+qr=(p+q)(p+r)$.

Proof. We shall prove the desired statements by constructing truth-tables.

(*a*)

TABLE 10.6
Truth table for p (q + r) = pq + pr

Row	p	q	r	q + r	p(q + r)	pq	pr	pq + pr
1	1	1	1	1	1	1	1	1
2	1	1	0	1	1	1	0	1
3	1	0	1	1	1	0	1	1
4	1	0	0	0	0	0	0	0
5	0	1	1	1	0	0	0	0
6	0	1	0	1	0	0	0	0
7	0	0	1	1	0	0	0	0
8	0	0	0	0	0	0	0	0

Since the columns headed by $p(q+r)$ and $pq+pr$ are identical, therefore we have

$$p(q+r)=pq+pr.$$

TABLE 10.7

Truth table for $p+qr=(p+q)(p+r)$

Row	p	q	r	qr	p + qr	p + q	p + r	(p + q) (p + r)
1	1	1	1	1	1	1	1	1
2	1	1	0	0	1	1	1	1
3	1	0	1	0	1	1	1	1
4	1	0	0	0	1	1	1	1
5	0	1	1	1	1	1	1	1
6	0	1	0	0	0	1	0	0
7	0	0	1	0	0	0	1	0
8	0	0	0	0	0	0	0	0

Since the entries in the columns headed by $p+qr$ and $(p+q)(p+r)$ are identical, therefore we have

$$p+qr=(p+q)(p+r).$$

Remark. If we use the symbols $\vee$ and $\wedge$ instead of $(+)$ an juxtaposition respectively, then the statements proved above can be written as follows :

$$\boxed{\begin{aligned} p \wedge (q \vee r)&=(p \wedge q) \vee (p \wedge r) \\ p \vee (q \wedge r)&=(p \vee q) \wedge (p \vee r) \end{aligned}}$$

Theorem 10·4. *For every statement p, there exists a statement p' (the negation of p) such that*

$$p+p'=1$$
$$pp'=0.$$

Proof. Follows immediately from table 10·4 and the meaning of 1 and 0 as statements.

10·5·3. Negation of a Compound Statement (De Morgan's Rules).

You are already familiar with De Morgan's rules for sets. Similar rules hold for statements as well.

Theorem 10·5. (*De Morgan's rules*). *If p, q be any statements, then*

(*a*) $(p+q)'=p'q'$

(*b*) $(pq)'=p'+q'$.

Proof. Let us construct truth tables for checking the truth of above statements.

(*a*)

TABLE 10.8
Truth table for (p + q)' = p'q'

Row	p	q	p + q	(p + q)'	p'	q'	p'q'
1	1	1	1	0	0	0	0
2	1	0	1	0	0	1	0
3	0	1	1	0	1	0	0
4	0	0	0	1	1	1	1

In the above table the entries in the columns headed by $(p+q)'$ and $p'q'$ are identical. Therefore we have

$$(p+q)'=p'q'.$$

(*b*)

TABLE 10.9
Truth table for (pq)' = p' + q'

Row	p	q	pq	(pq)'	p'	q'	p' + q'
1	1	1	1	0	0	0	0
2	1	0	0	1	0	1	1
3	0	1	0	1	1	0	1
4	0	0	0	1	1	1	1

In the above table the entries in the columns headed by $(pq)'$ and $p'+q'$ are identical. Therefore we have

$$(pq)'=p'+q'.$$

Remark. If we use $\sim$, $\vee$, $\wedge$ instead of '/', +, and juxtaposition respectively, then De Morgan's rules can be expressed in the following form :

$$\boxed{\begin{aligned}\sim (p \vee q) &= (\sim p) \wedge (\sim q)\\ \sim (p \wedge q) &= (\sim p) \vee (\sim q)\end{aligned}}$$

10·5·4. Algebra of statements

As a consequence of theorems proved above we often talk of the Boolean algebra of statements. More precisely, we have the following definition :

Definition 10·2. *A class B of statements together with two binary operations* (+) *and* (.) *(where a.b is written as ab) is said to be a Boolean algebra if the following hold :*

P1. B is closed for the operations (+) and (.).

P2. The operations (+) and (.) are commutative.

P3. There exist in B distinct identity elements 0 and 1 relative to the operations (+) and (.) respectively.

P4. Each operation is distributive over the other.

P5. For every p in B there exists p' in B such that

$$p+p'=1,$$

$$pp'=0.$$

Remark. In view of theorems 10·1 to 10·4 in order to check that a class of B of statements is a Boolean algebra, it is enough to check (*i*) P1, , (*ii*) 0 and 1 are in B, (*iii*) if p is in B, then p' is also in B.

10·5·5. Duality

Observe that theorems 10·1 to 10·4 are symmetric with respect to the operations (+) and (.), and the two identities 0 and 1. An important consequence of this fact is the following **Principle of Duality for Statements** :

Every assertion or algebraic identity involving statements remains valid if the operations (+) *and* (.), *and the identity elements 0 and 1 are interchanged throughout.*

The two De Morgan's rules proved in Theorem 10·5 are duals of each other in the sense that each of them can be obtained from the other by applying the principle of duality.

EXERCISE 10 (*c*)

1. If p stands for the statement 'It is hot', and q stands for the statement 'It is blowing', then state in words the following :

(*a*) $p+q$ (*b*) pq

(*c*) pq' (*d*) $p'q$.

2. Using suitable symbols express the following as conjunctions or disjunctions :

   (*a*) He must stop or he will faint.

   (*b*) The sky is blue and the grass is green.

   (*c*) He is intelligent and she is beautiful.

   (*d*) Ice is cold or 10 is a prime.

3. Write each of the following statements in symbolic form using p and q, where p : He is intelligent, q : He is ugly.

   (*i*) He is intelligent and ugly.

   (*ii*) He is intelligent but not ugly.

   (*iii*) He is neither intelligent nor ugly.

   (*iv*) He is intelligent or he is unintelligent and ugly.

   (*v*) It is not true that he is intelligent or ugly.

   (*vi*) He is not intelligent or ugly.

4. Classify the following statements as true for all statements p, q, and r; false for all statements p, q, and r or sometimes true and sometimes false ;

   (*a*) $p+p'$ (*b*) pp'

   (*c*) $pq(r+r')$ (*d*) $(p+q)rr'$

   (*e*) $p+(q+r)$ (*f*) $pp'+(qq'+rr')$

   (*g*) $[(p+q)(p+q')]\,[(p'+q)(p'+q')]$.

5. Let p, q and r be statements such that p is false, q is true, and r is true. Decide whether each of the following is true or false :

   (*a*) p' (*b*) q'

   (*c*) pq' (*d*) $q'r$

   (*e*) $p'r'$ (*f*) $(pq)r'$

   (*g*) $p'(qr')'$ (*h*) $p(q'r)$.

10·6. BOOLEAN ALGEBRA

Consider a set $S=\{0, 1\}$, consisting of two elements denoted by the symbols 0 and 1. We define two binary operations on S, to be called the *logical sum* and *logical product*, and denoted by the symbols '+' and '.' respectively, by means of the following operation tables :

TABLE 10.10

+	0	1
0	0	1
1	1	1

·	0	1
0	0	0
1	0	1

It can be easily verified that the triple (S, +, .) satisfies the following properties :

B 1. *Closure.* S is closed for the operations '+' and '.'.

B 2. *Commutativity.* The operations '+' and '.' are commutative.

B 3. *Identities.* There exist in S distinct identity elements **0** and 1 relative to the operations '+' and '.' respectively.

B 4. *Distributivity.* Each operation is distributive with respect to the other.

B 5. *Complements.* For each x in S, there exists an element x' in S such that

$$x+x'=1,\ xx'=0$$

Because of the above properties we say that (S, +, .) is a Boolean algebra. It is, in fact, the simplest example of a Boolean algebra, which we define below :

Definition 10·3. *A set S consisting of atleast two elements is said to be a Boolean algebra for two binary operations '+' and '.' defined on it if the properties* **B1—B5** *(listed above) are satisfied.*

Boolean algebra has important applications to the theory of switching circuits as also to computer design. It is because of these applications that more and more mathematicians and engineers are attracted these days towards Boolean algebra.

Illustrations. 1. As already remarked above, a class B of statements satisfying certain conditions already stated provides examples of Boolean algebra for the two operations 'disjunction' and 'conjunction'.

2. If S be any non-empty set, then the set P (S) of subsets of S is a Boolean algebra for the operations 'union of sets' ($\cup$) and 'intersection of sets' ($\cap$). The elements ϕ and S serve as the identities for $\cup$ and $\cap$ respectively. For each element A of P (S), S ~ A is the element of P(S) such that

$$A\cup(S\sim A)=S, \text{ and } A\cap(S\sim A)=\phi.$$

Remark. The symbols 0 and 1 in the example considered in the beginning of this section, as also in the definition of a Boolean

algebra, ought not to be confused with the numbers 0 and 1. They have no relationship with the numbers 0 and 1. We could have used any other symbols as well. Similarly, the symbols '+' and '.' used here for logical addition and logical multiplication have no relationship with addition and multiplication of numbers. We could have equally well used $\cup$ and $\cap$, or $\vee$ and $\wedge$, or any other symbols. The symbols 0, 1, '+', '.' have been used just because of convenience.

It can be easily shown that in a Boolean algebra, the following properties also hold :

B 6. *Idempotence.* For every element x in a Boolean algebra,

$$x+x=x, \ xx=x.$$

B 7. For each element x in a Boolean algebra

$$x+1=1, \ x \,.\, 0=0$$

B 8. *Absorption.* For each pair of elements x, y in a Booleen algebra.

$$x+xy=x, \ x(x+y)=a.$$

B 9. *Associativity.* In every Boolean algebra, each of the binary operations (+) and (.) is associative, that is, for every x, y and z in S,

$$x+(y+z)=(x+y)+z \text{ and } x(yz)=(xy)z.$$

B 10. *De Morgan's Rules.* For every x and y in a Boolean algebra,

$$(xy)'=x'+y' \text{ and } (x+y)'=x'y.'$$

B 11. *Complementation.* In any Boolean algebra,

$$0'=1 \text{ and } 1'=0.$$

B 12. *Involution.* For every x in a Boolean algebra,

$$(x')'=x.$$

10·6·1. Principle of duality

You must have observed that in all the properties B 1—B 12 listed above, there is a certain symmetry with respect to '+', '.', '0' and '1' in the sense that if in any statement '+' and '.' are in interchanged and the two identities '0' and '1' are interchanged, then the resulting statement is also a valid statement. As a consequence of this symmetry, we have the following important principle :

Theorem 10·6. (*Principle of duality*). *Every statement or identity deducible from the axioms of a Boolean algebra remains valid if the two operations '+' and '.' are interchanged, and the two identities 0 and 1 are interchanged.*

If a statement or an identity is deduced from a given statement or identity by applying the principle of duality once, then it is said to be the dual of the given statement. It is obvious that the dual of a statement is the original statement itself.

10·6·2. Boolean function

Let (S, +, .) be a Boolean algebra, with identities 0 and 1. A function $f : S \to S$ is said to be a *Boolean function* if x takes only the value 0 or 1. In other words, a Boolean function f is a function from S to {0, 1}.

Illustrations. $f : (x, y, z)\, f(x) \to x+y'+z,$

$g : (x, y, z)\, f(x) \to x'yz+xyz'+x'y'z$

$h : (x, y, z)\, f(x) \to (x+y)(y+z)+xy$

are all Boolean functions.

Since f, g, h in the above illustrations are completely determined by the expressions

$$x+y'+z,\ x'yz+xyz'+x'y'z,\ (x+y)(y+z)+xy$$

respectively, therefore these expressions are also sometimes referred to as Boolean functions. Strictly speaking, they are all Boolean expressions. We shall use the term Boolean expression to mean a Boolean function whenever there is no chance of ambiguity.

EXERCISE 10 (*d*)

1. Verify all the properties B1—B12 for the Boolean algebra described by the operations in Table 10·10 on {0, 1}.
2. Verify all the properties B1—B12 for the algebra of statements.
3. Deduce properties B6—B12 from B1—B5 for a Boolean algebra.
4. Prove that for any x, y and z in a Boolean algebra, the following four expressions are equal :

   (*a*) $(z+x)(z'+y)(x+y)$ (*b*) $yz+xz'+xy$

   (*c*) $yz+xz'$ (*d*) $(z+x)(z'+y)$.
5. Show that no Boolean algebra can have exactly three distinct elements.

6 Show that the set $\{a, b, c, d\}$ with operations (+,) defined below, is a Boolean algebra.

+	a	b	c	d
a	a	b	c	d
b	b	b	b	b
c	c	b	c	b
d	d	b	b	d

	a	b	c	d
a	a	a	a	a
b	a	b	c	d
c	a	c	c	a
d	a	d	a	d

10·7. CONDITIONAL AND BICONDITIONAL STATEMENTS

Consider the following statements :

(*i*) In a triangle ABC,

if $\angle A$ is an obtuse angle, then

$AB^2+AC^2>BC^2$.

(*ii*) In a triangle ABC,

if $\angle A$ is an acute angle, then

$AB^2+AC^2=BC^2$.

Both the above statements are of the form : If p, then q. A statement of the form 'If p, then q' is called a **conditional statement** or an **implication**. In symbols, it is written as $p \rightarrow q$. We have already constructed truth tables for several compound statements. We would naturally like to construct a truth table for $p \rightarrow q$ as well. Before we do that, let us consider the following statements :

(*i*) If 6 is an even number, than 7 is an odd number.

(*ii*) If 3 is an odd number, then 49 is a perfect square.

(*iii*) If 5 is a prime, then 21 is a prime.

(*iv*) If 8 is an odd number, then 23 is a prime.

(*v*) If 16 is a prime, then 81 is a prime.

Each of the above statements is of the type $p \rightarrow q$ for some statements p and q.

For (*i*), let us write

p=6 is an even number,

q=7 is an odd number.

Both p and q are true, and q can be logically deduced from p by using the fact that even and odd numbers occur alternately.

According to common sense, $p \rightarrow q$ should have the truth value 1.

For (*ii*), let us write

p=3 is an odd number,

q=49 is a perfect square.

Both p and q are true statements, but q cannot be logically deduced from p.

According to common sense $p \rightarrow q$ should be assigned the truth value 0.

For (*iii*), let us write

p=5 is a prime,

q=21 is a prime.

The statement p is known to be true, and the statement q is known to be false.

According to common sense, $p \to q$ should have the truth value 0.

For (*iv*), let us write

p=8 is an odd number,

q=23 is a prime.

Here p is false and q is true. It should not be possible to deduce a true statement from a false statement. Therefore according to common sense, $p \to q$ ought to be given the truth value 0.

For (*v*) let us write

p=16 is a prime

q=81 is a prime.

Both p and q are false, and according to common sense $p \to q$ should have the truth value 0.

Let us tabulate the truth values of p and q in the above five cases, alongwith the possible truth values that $p \to q$ should have according to our common sense

	p	q	$p \to q$
(*i*)	1	1	1
(*ii*)	1	1	0
(*iii*)	1	0	0
(*iv*)	0	1	0
(*v*)	0	0	0

The above assignments of truth values is on the assumption that there should be causal relationship between p and q, and dictates of common sense should be honoured. In mathematical logic we do not do this. For, if we were to do so, we would be in difficulty. In cases (*i*) and (*ii*), there would be an immediate contradiction. In both the cases (*i*) and (*ii*), both p and q have the truth-value 1, but while in case (*i*) $p \to q$ should be given the truth-value 1 and in case (*ii*) $p \to q$ should be given the truth value 0. How do we assign truth values to $p \to q$ in such a situation? This difficulty is overcome by the following definition :

Definition 10·4. *If p and q are any statements, then*

$$p \to q = p' + q.$$

In the above definition, p is called the antecedent and q is called the consequent.

Remark. The implication $p \to q$ does not mean that p and q are related in such a way that q can be derived from p Let us once again emphasize that there need not be any causal relation between p and q.

By using the definition of $p \to q$ we can easily construct the truth table for $p \to q$ as shown in Table 10·11.

TABLE 10.11
Truth table for p → q

Row	p	q	p → q
1	1	1	1
2	1	0	0
3	0	1	1
4	0	0	1

Biconditional Statement. If p and q be any two statements, then the relation '$\leftrightarrow$' is defined by the equation $p \leftrightarrow q = (p \to q)\ (q \to p)$. It is called a *biconditional statement* or *equivalence.* It is read as 'p if and only if q'. It has the truth value 1 in exactly those cases in which the truth values of p and q are the same. Table 10·12 gives the truth-table for the biconditional statement.

TABLE 10.12
Truth tablo for p ↔ q

Row	p	q	p ↔ q
1	1	1	1
2	1	0	0
3	0	1	0
4	0	0	1

Remarks 1. Given an implication $p \to q$ we frequently talk of three other implications :

(*a*) The *converse* of $p \to q$. It is the implication $q \to p$.
(*b*) The *inverse* of $p \to q$. It is the implication $p' \to q'$.
(*c*) The *contrapositive* of $p \to q$. It is the implication $q' \to p'$.

The following table compares the truth values for the above four implications connecting two given propositions p and q.

TABLE 10.13

Row	p	q	$p \to q$	$q \to p$	$p' \to q'$	$q' \to p'$
1	1	1	1	1	1	1
2	1	0	0	1	1	0
3	0	1	1	0	0	1
4	0	0	1	1	1	1

From the above table we find that $p \to q$ and $q' \to p'$ are equal, *i.e.*, the original implication and its contrapositive are equal. Also

TABLE 10.14

S.No.	Statement	Symbolic Notation
1.	If p then q	$p \to q$
2.	p only if q	$p \to q$
3.	p is a sufficient condition for q	$p \to q$
4.	A necessary condition for p is q	$p \to q$
5.	In order that p it is necessary that q	$p \to q$
6.	p unless q	$q \to p$
7.	p if q	$q \to p$
8.	p is a necessary conditon for q	$q \to p$
9.	In order that p it is sufficient that q	$q \to p$
10.	A sufficient condition for p is q	$q \to p$
11.	p if only if q	$p \leftrightarrow q$
12.	p is a necessary and sufficient condition for q	$p \leftrightarrow q$

we find that $q \to p$ and $p' \to q'$ are equal, *i.e.*, the inverse and converse of an implication are equal.

2. From the definition of $p \to q$ and from De Morgan's law, we find that the negation of $p \to q$ is given by $(p \to q)' = (p' + q)' = pq'$. In other words, it is false that p implies q' can also be stated by saying 'p and not q'.

3. The following table gives several abbreviated statements and their translations into symbolic notations. These will serve as definitions of the connectives commonly used in the English language and will be found useful in translating assertions about statements into symbolic language.

Example 6. *Let p be the statement '5 is an odd integer' and let q be the statement 'chocolate is sweet'. Write in words*

(*a*) the implication $p \to q$ (*b*) its converse (*c*) its inverse (*d*) its contrapositive and (*e*) its negation.

Solution. (*a*) If 5 is an odd integer, then chocolate is sweet.

(*b*) If chocolate is sweet, then 5 is an odd integer.

(*c*) If 5 is an odd integer, then chocolate is not sweet.

(*d*) If chocolate is not sweet, then 5 is an odd integer.

(*e*) 5 is an odd integer, and chocolate is not sweet.

Example 7. *Label suitable simple statements as p and q and translate the following into symbols* :

(*a*) If sugar is cheap, then sweets are served in plenty.

(*b*) Sweets are served in plenty, unless sugar is cheap.

(*c*) A necessary condition for sugar to be cheap is that sweets are rarely served.

(*d*) Sweets are rarely served only if sugar is costly.

(*e*) Sugar is cheap and sweets are rarely served.

Solution. Let p=sugar is sweet, q=sweets are served in plenty. Then p'=sugar is costly, q'=sweets are rarely served. The symbolic translations of the given propositions are :

(*a*) $p \to q$ (*b*) $p' \to q$ (*c*) $p \to q'$ (*d*) $q' \to p'$
(*e*) pq'.

Example 8. *Show that if 'p' and 'q' be any two statements, then the statement* $(p\,(p \to q)) \to q$ *is a tautology*.

Solution. $(p\,(p \to q)) \to q$

$$= (p\,(p' + q)) \to q,$$
$$= (pp' + pq) \to q,$$
$$= (0 + pq) \to q,$$

$$=pq\to q,$$
$$=(pq)'+q,$$
$$=p'+q'+q,$$
$$=p'+1,$$
$$=1.$$

Hence $(p\ (p\to q))\to q$ is a tautology.

Example 9. *Show, by constructing a truth table that the implication*

$$[(p\to q)\ (q\to r)]\to(p\to r)$$

is a tautology.

Solution.

TABLE 10.15

Row	p	q	r	p → q	q → r	(p → q) (q → r)	p → r	[(p → q)(q → r) → (p → r)
1	1	1	1	1	1	1	1	1
2	1	1	0	1	0	0	0	1
3	1	0	1	0	1	0	1	1
4	1	0	0	0	1	0	0	1
5	0	1	1	1	1	1	1	1
6	0	1	0	1	0	0	1	1
7	0	0	1	1	1	1	1	1
8	0	0	0	1	1	1	1	1

Since all the entries in the last column are '1', therefore the given proposition is a tautology.

EXERCISE 10 (*e*)

1. Let p=The dog barks, q=The dog is scared. Write each of the following statements in symbolic form :

   (*a*) The dog is scared only if it barks.

   (*b*) The dog never barks when it is scared.

   (*c*) The dog barks only if it is scared.

(*d*) The dog is scared only if it barks.
(*e*) The dog barks if and only if it is scared.
(*f*) The dog never barks if it is not scared.

2. Write down the truth values of the following statements :
   (*a*) If 10 is an even number, then 25 is a perfect square.
   (*b*) If 16 is an odd number, then 37 is a prime.
   (*c*) If 101 is a prime, then 100 is a perfect square.
   (*d*) If 20 is a perfect square, then 16 is a perfect cube.
   (*e*) If 8 is a perfect cube, then 17 is an odd number.
   (*f*) If 6 is a composite number, then 13 is a composite number.

3. Write in reasonable English the negation of each of the following statements :
   (*a*) He will die unless he is given blood.
   (*b*) A necessary condition for two triangles to be congruent is that they are equiangular.
   (*c*) I grow fat only if I eat ice-cream.

4. Write in words the converse, inverse, contrapositive, and negation of the implication 'If 6 is greater than 5, then $\frac{1}{5}$ is greater than $\frac{1}{6}$'.

5. Show, by constructing, truth tables that each of the following implications is a tautology ;
   (*a*) $pq \to pq$ (*b*) $pq \to p$
   (*c*) $p' \to (p \to q)$ (*d*) $(p+q)\, p' \to q$
   (*e*) $p \to (p+q)$ (*f*) $q \to (p \to q)$.

10.8. VALID ARGUMENTS

In mathematics we often assume a certain sets of propositions without proof, and deduce from them by logical reasoning, certain other propositions. The propositions that are assumed without proofs are called *axioms*. We are not concerned with the absolute truths of the propositions that we deduce. What we are concerned is whether we can logically derive a proposition from the axioms. As an illustration, consider the pythagorean proposition which says that "In a right-angled triangle the sum of the squares of the sides containing the right-angle is equal to the square on the hypotenuse". This proposition is true for right-angled triangles in a plane because it can be deduced from the axioms of Euclidean plane geometry.§ It is however, not true for triangles on the surface of a sphere.

In the present section we shall study the processes by which we can derive a proposition (called **conclusion**) from a set of given propositions (called **premises** or **hypotheses**). A process by which a conclusion is formed from given premises is called an **argument**. An argument is said to be **valid** if and only if the conjunction of the premises implies the conclusion.

Definition 10.5. *An argument which yields a conclusion r from premises $p_1, p_2, \ldots, p_n$ is said to be valid if and only if the proposition $(p_1 p_2 \ldots p_n) r$ is a tautology.*

How to check the validity of an argument ? There are three general methods for checking the validity of a given argument

Method I. The validity of a given argument can be checked directly from the definition by using a truth table. That is, we can construct a truth table for the proposition

$$(p_1 p_2 p_3 \ldots p_n) \rightarrow r. \qquad \ldots(A)$$

If we find from the truth-table that the above proposition is a tautology, then the argument is valid, otherwise it is not valid.

Method II. We can simplify the proposition (A) by standard methods of simplification. If the proposition reduces to 1, then the argument is valid, otherwise it is not valid.

Method III. The third method for checking the validity of an argument is to reduce the argument to a chain of arguments, each of which is known to be valid (as a consequence of previous checking).

10.8.1. Some Important Valid Arguments

The following two arguments are the most frequently used valid arguments :

(1) *Rule of detachment.* The rule of detachment states that the propositions 'p' and '$p \rightarrow q$' together yield the proposition 'q'. It is also called *modus ponens.* The rule of detachment is often written as follows :

$$\begin{array}{c} p \\ p \rightarrow q \\ \hline q \end{array}$$

The above schematic arrangement consists of the following :

(*a*) The premise or premises are listed first.

(*b*) A horizontal line is drawn when all the premises have been listed.

(*c*) The conclusion is written below the horizontal line.

(*d*) Reason or explanations, if any, are written to the right of each proposition.

The validity of the rule of detachment has already been verified in example 8 by using method II. We can also verify it by method I.

(2) *Law of Syllogism.* The law of syllogism is given by the form

$$\begin{array}{c} p \rightarrow q \\ q \rightarrow r \\ \hline p \rightarrow r \end{array}$$

The validity of the above argument has already been verified in example 9 by using method I. We can also verify it by using method II.

In addition to the above two forms of valid argument—the rule of detachment and the law of syllogism, the following six forms of valid arguments are also useful. Their validity has been checked in problem 5 of exercise 10 (*e*) earlier.

TABLE 10.16
Forms of Valid Argument

Form 1	Form 2	Form 3	From 4	From 5	From 6
$\begin{array}{c} p \\ q \\ \hline pq \end{array}$	$\begin{array}{c} pq \\ \hline p \end{array}$	$\begin{array}{c} p' \\ \hline p \rightarrow q \end{array}$	$\begin{array}{c} p+q \\ p' \\ \hline q \end{array}$	$\begin{array}{c} p \\ \hline p+q \end{array}$	$\begin{array}{c} q \\ \hline p \rightarrow q \end{array}$

While checking the validity of an argument, the following rule of substitution is often used :

Rule. A valid argument remains valid if *any* occurance of a proposition is replaced by an equivalent proposition.

Remarks 1. The validity of otherwise of an argument is independent of the truth or falsity of the conclusion. That is, an argument may be valid but the conclusion may be false. Also, an argument may be invalid but the conclusion may be true. For example, consider the following examples :

(A) Ice is hot.

<u>If ice is hot, then snow is green.</u>

Snow is green.

The above agrument is valid but the conclusion 'snow is green, is false.

(B) 6 is an even integer.

<u>If 5 is an odd integer, then 6 is an even integer.</u>

5 is an odd integer,

Although the conclusion '5 is an odd integer' is true, the argument used above is false.

2. While checking an argument for validity, if we find (or even suspect !) that the argument is invalid, the proof of invalidity can be given simply by exhibiting a set of truth values of the various propositions involved for which the premises are all true but the conclusion is false. This would show that if we were to construct the truth table, one row in the truth table would contain a 0, and consequently the argument is not valid.

Example 10. *Examine whether the following argument is valid*

$$\begin{array}{c} p \to q \\ q \to r \\ p \\ \hline r \end{array}$$

Solution 1. Let us write

$$s = [(p \to q)\ (q \to r)\ p] \to r.$$

We construct the truth table for the proposition s.

Since the s column contains only 1's, the argument is valid.

TABLE 10.17

Row	p	q	r	p → q	q → r	(p → q) (q → r)p	s
1	1	1	1	1	1	1	1
2	1	1	0	1	0	0	1
3	1	0	1	0	1	0	1
4	1	0	0	0	1	0	1
5	0	1	1	1	1	0	1
6	0	1	0	1	0	0	1
7	0	0	1	1	1	0	1
8	0	0	0	1	1	0	1

Solution 2.

$$\begin{aligned} s &= [(p \to q)\ (q \to r)\ p] \to r, \\ &= [(p'+q)\ (q'+r)\ p]'+r, \\ &= pq'+qr'+p'+r, \\ &= p'+pq'+qr'+r, \end{aligned}$$

$$=(p'+p)\,(p'+q')+qr'+r,$$
$$=1(p'+q')+(q+r)(r+r'),$$
$$=p'+q'+(q+r)1,$$
$$=p'+q'+q+r,$$
$$=p'+1+r,$$
$$=1.$$

Sinces reduces to 1, the argument is valid.

Solution 3. Consider the following chain of arguments :

$p\to q$	a premise
$q\to r$	a premise
$p\to r$	law of syllogism
p	a premise
r	rule of detachment

Observe that the above chain of valid arguments gives a neater method of showing the validity of the given argument as compared to Methods I and II.

10·8·2. Some Applications of Logic in Solving Problems

Example 11. *Anupama, Bhavna, Chitra, and Deepti competed for a scholarship. Someone asked them, "What was the result ?" Anupama said, "Chitra topped, Bhavna was second." Bhavna said, "No, Chitra was second and Deepti was third." Chitra said, "Deepti was last, Anupama was second." Each of three girls made three assertions, of which only one was true. Who won the scholarship ?*

Solution. In this problem each of the three girls makes two statements, and exactly one of the two statements is known to be true.

If p and q are two propositions, and exactly one of them is false, then we know that

$$p+q=1, \qquad pq=0.$$

Neither of the above relations by itself is sufficient, because the first one is true even when both p and q are true and the second one is true even when both p and q are false. Let us try to search for a single relation which is true when exactly one of the two propositions p and q is true and the other is false. This could happen in two ways : either p is true and q is false, or, p is false and q is true. That is, either $pq'=1$ or $p'q=1$, but not both. This means that

$$pq'+p'q=1 \text{ and } (pq')\,(p'q)=0.$$

Now $(pq')\,(p'q)$ is always 0, Therefore the given condition is equivalent to the relation $pq'+p'q=1$. We shall use this relation

to symbolize all the three statments. Let A_1 be the proposition 'Anupama was first', A_2 the proposition 'Anupama was second', and so on—each letter indicating the person about whom the assertion was made, and the numeral indicating the ranking of the person about whom the assertion is made. For example C_3 would stand for the asseration that Chitra was third. Let us now express the three statements in symbolic form.

Anupama's statement : $C_1B_2'+C'_1B_2=1$.

Bhavana's statement : $C_2D_3'+C'_2D_3=1$.

Chitra's statement : $D_4A_2'+D_4'A_2=1$.

Since the conjunction of true propositions is a true proposition, therefore writing the conjuction of all the above statements we have

$$(C_1B_2'+C_1'B_2)\,(C_2D'_3+C_2'D_3)(D_4A'_2+D_4'A_2)=1.$$

Since distributive law holds, we can multiply out, and write the left hand side as the sum of eight term. We thus have

$$\begin{aligned}&C_1B_2'C_2D_3'C_4A_2'+C_1B_2'C_2'D_3'A_2\\&+C_1B_2'C_2'D_3D_4A_2'+C_1B_2'C_2'D_3'D_4'A_2\\&+C_1'B_2C_2D'_3D_4A_2'+C_1'B_2C_2'D_3D_4'A_2\\&+C_1'B_2C_2'D_3D_4A'_2+C_1'B_2C_2D_3'D'_4A_2\\&=1.\end{aligned}$$

The first term on the left contains C_1C_2 which is zero because Chitra could not have come first as well as second. The third, fourth and seventh terms are zero for a similar reason. The fifth term contains B_2C_2 which is zero because both Bhavana and Chitra could not have come second. Therefore the flfth term is zero. The sixth and the eighth terms ars also zero for similar reasons. We are thus left with the second term only. That is, the above relation yields

$$C_1B_2'C'_2D_3D_4'A_2=1,$$

which gives $C_1=B_2'=C_2'=D_3=D_4'=A_2=1$.

Now $C_1=1$ means Chitra was first,

$A_2=1$ means Anupama was second,

$D_3=1$ means Deepti was third,

and consequently Bhavana was fourth.

Example 12. *Consider the following statements :*

(1) *All writers who understand human nature are clever.*

(2) *No one is a true poet unless he can stir the hearts of men.*

(3) *Shakespeare wrote Hamlet.*

(4) *No writer who does not understand human nature can stir the hearts of men.*

(5) *None but a true poet could have written Hamlet.*
What conclusion can you draw form the above assertions ?

Solution. Let us first of all note that the universal class here is the class of "writers". Let us denote it by w.

We shall denote by a the class of writers who understand human nature, by b the cless of clever writers, by c the class of poets, by d the class of writers who can stir the hearts of men, by s the class consisting of the single writer, namely Shakespeare, and by h the class consisting of the writer of Hamlet. Let us now put all the given assertions in symbolic form

(1) can be written as $a \subset b$.
(2) can be written as $c \subset d$.
(3) can be written as $s = h$.
(4) can be written as $a' \subset d'$.
(5) can be written as $h \subset c$.

Observe that $a' \subset d'$ is equivalent to the statement $d \subset a$. From the have five relations we have

$$s = h \subset c \subset d \subset a \subset b.$$

This leads us to the donclusion that $s \subset b$, *i.e.*, Shakespeare was a clever writer.

Exercise 10 (f)

1. Check the validity of each of the following arguments :

(a) $\begin{array}{c} q' \\ \underline{p \to y} \\ p' \end{array}$ (b) $\begin{array}{c} pq \\ \underline{p' \to q} \\ q' \end{array}$

(c) $\begin{array}{c} p \leftrightarrow q' \\ q \leftrightarrow r \\ \underline{r} \\ p \end{array}$ (d) $\begin{array}{c} p \to q \\ \\ \underline{r \to q'} \\ p \to r' \end{array}$

2. Show that the following argument is valid :

$$\begin{array}{c} q \\ \underline{q \to r} \\ p \to r \end{array}$$

3. Is the following argument valid :

$$\begin{array}{c} p \\ q' + r \\ \underline{p' \to q} \\ r \end{array}$$

Give reasons for your answer.

4. Prove that the following arguments are valid :

(a)
$$\begin{array}{l} p \to q \\ p + s \\ s' \\ \hline p \end{array}$$

(b)
$$\begin{array}{l} p' \\ p + s \to pr' \\ \hline s' \end{array}$$

5. We are given the following statements, as premises, all of which refer to a certain meal :

(a) If I take tea, I don't drink milk.

(b) I eat cream wafers only if I drink milk.

(c) I don't take apple unless I eat cream wafers.

(d) In the evening to-day, I had tea.

Is it possible to draw any conclusion about whether I took an apple to-day in the evening.

If so, what is the correct conclusion ?

10·9. SWITCHING CIRCUITS

One of the important applications of Boolean algebra is to switching circuits. A switch is a two-state device, the two states being closed and open. The algebra of circuits is of interest to mathematicians as well as to engineers. The interest of the mathematicians arises from the fact that Boolean algebra is used in the design and simplification of complex circuits involved.

We shall denote a switch by a single lower case letter a, b, c, x, y, z, ... If two switches operate in such a manner that they always open and close simultaneously, *i.e.*, if one of them is open, then the other is also open, and if one of them is closed, then the other is also closed, then we shall denote them by the same letter. If two switches operate in such a manner that whenever the first is closed, the second is open ; and whenever the first is open, the second is closed. We denote them in such a manner that if the first is denoted by x, the second is denoted by x' (or if the second is denoted by x, then the first is denoted by x').

A switch which is always closed is represented by 1 and a switch which is always open is represented by 0.

Whatever we have said about switches applies to circuits as well. If a circuit is closed, so that current passes through it, we say that it is in state 1 ; if a circuit is open so that current does not pas through it, it is said to be in state 0.

10·9·1. Connection of Two Switches in Parallel

If two swithes x and y are connected in parallel, as in Fig. 10·3(a) (so that current passes through the circuit if one of the

switches is closed), then the resulting circuit is denoted by $x+y$.

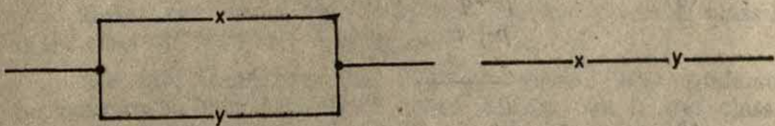


Fig. 10·3 (a) Fig. 10·3 (b)

10·9·2. Connection of Two Switches in Series

If two switches x and y are connected in series as in Fig. 10·3 (*b*) (so that current passes through the circuit only if both the switches are closed), then the resulting circuit is denoted by xy.

From the above description we find that to each series-parallel circuit there corresponds an algebraic expression. Conversely, it is obvious that to each algebraic expression involving the variables x, y, z, etc., and the symbols $+$,., and $'$ there corresponds a series-parallel circuit having switches $x, y, z, \ldots\ldots, x', y', z', \ldots\ldots$ etc.

We express the above relationship between algebraic expressions (functions) and circuits by saying that the function represents the circuit and the circuit realizes the function.

10·9·3. Equivalent Circuits

Two circuits involving switches $x, y, \ldots$ are said to be equivalent if the closure conditions of the circuits are the same for any given position of the switches involved. This means that two circuits are equivalent if for every position of the switches, either current passes through both the circuits, or it does not pass through either circuit.

Two algebraic expressions are said to be equal if and only if they represent equivalent circuits.

Since a switch can be in one of the two states 0 and 1, and algebraic expressions represents circuits here, therefore each of the variables in the expressions to be encountered in the algebra of switching circuits can take two values 0 and 1, therefore it is very natural to ask as to whether the boolean algebra on the two-elementic-set $\{0, 1\}$ that we described earlier—can be used to assign a value to an algebraic expression for given values of the variables involved. The important question is : Will this assignment serve any useful purpose for switching circuits.? Fortunately the answer is that it indeed serves the purpose. The correspondence between series/parallel circuits and the expressions $x+y$, xy has been made in that way. The situation will become clear by having a glance at the following table of closure properties of switching functions (as the algebraic expressions that correspond to circuits are sometimes called) x', $x+y$ and xy.

TABLE 10.18
Closure properties of switching functions x′, x + y, xy

Row	X	Y	X′	X + Y	XY
1	1	1	0	1	1
2	1	0	0	1	0
3	0	1	1	1	0
4	0	0	1	0	0

Let us compare this table with the corresponding truth tables for statements. Are you not surprised that both are the same ? Let us compare it once again, this time with the corresponding tables for the '+' and ' . ' for Boolean algebra (page 580). Both turn out to be the same. That is, the truth tables for disjunction and conjunction of statements, the operation tables for logical sum and logical product in a Boolean algebra on a set consisting of two elements, and tables of closure properties of switching functions, all of them are the same. That is the beauty of mathematics—the same structure appearing in widely differing situations. Now you must have understood as to why Boolean algebra should be applicable to switching circuits.

While the validity of the laws of Boolean algebra for switching circuits is an immediate consequence of the fact that, as we have seen above, the operation tables and the tables of closures properties for switching functions are identical, a direct verification is also possible by considering the circuits that realize the functions on either side of an identity such as

$$x(y+z)=xy+xz,$$

or
$$x+yz=(x+y)(x+z),$$

and examining whether the circuits turn out to be equivalent.

Example 13. *Draw the circuits which realize the functions $x+yz$ and $(x+y)(x+z)$. Are the circuits equivalent ? If yes, what conclusion can you draw ?*

Solution. The circuits are as shown in Fig. 10.4.

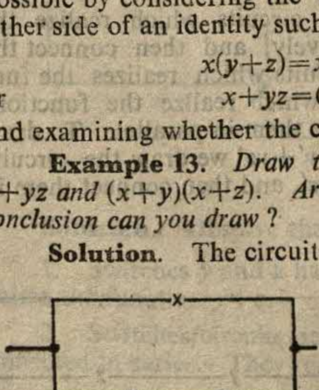


Fig. 10·4(a)

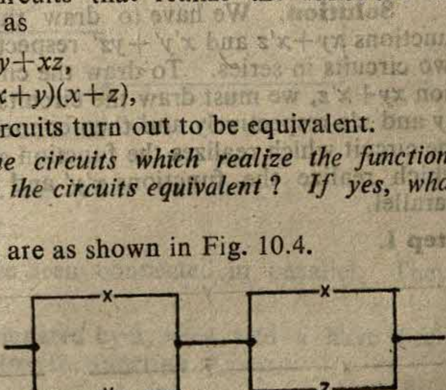


Fig. 10·4(b)

The circuit in Fig. 10·4(*a*) is closed if and only if the either switch x is closed, or both y and z are closed. The same is true of the circuit in Fig. 10.4(*b*). Therefore the two circuits are equivalent.

Conclusion. The distributive law

$$x+yz=(x+y)(x+z)$$

holds for switching circuits.

10·9·4. How to Apply Boolean Algebra to Switching Circuits ?

Before we proceed further, let us see as to what are the various problems in applying our knowledge of Boolean algebra to switching circuits. The main problems are the following :

1. Given a boolean function, to draw the switching circuit which realizes the function.

2. Given a switching circuit, to find the boolean function that represents the circuit.

3. Given a switching circuit, to obtain a simpler equivalent circuit.

So far as (3) is concerned, it consists of the following steps :

(*i*) to find the boolean function that represents the circuit.

(*ii*) to simplify the function by applying the laws of boolean algebra.

(*iii*) to draw a circuit which realizes the simplified function.

We shall take up the above problems (1)—(3) one by one. The examples given below will illustrate the method.

Example 14. *Draw a circuit which realizes the following fuuction :*

$$(xy+x'z)(x'y'+yz').$$

Solution. We have to draw the circuits which realize the functions $xy+x'z$ and $x'y'+yz'$ respectively, and then connect the two circuits in series. To draw the circuits which realizes the funtion $xy+x'z$, we must draw the circuits which realize the functions xy and $x'z$ respectively and then connect them in parallel. To draw the circuit which realizes the function $x'y'+yz'$ we draw the circuits which realize the functions $x'y'$ and yz', and then connect them in parallel.

Step I.

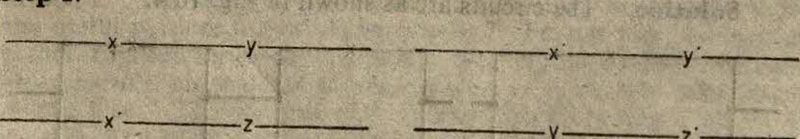


Fig. 10·5(a)

Step II.

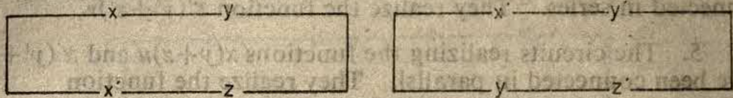


Fig. 10·5(b)

Step III.

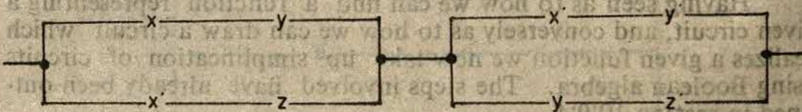


Fig. 10·5(c)

In step I we have drawn circuits for xy, $x'z$, $x'y'$, and yz' (Fig. 10·5(*a*)].

In step II we have connected the circuits as explained above to get the circuits for the functions $xy+x'z$ and $x'y'+yz'$ [Fig. 10·5(*b*)]. In step III we have connected the circuits for $xy+x'z$ and $x'y'+yz'$ in series to get the required circuit ag shown in Fig. 10·5(*c*).

Example 15. *Find the boolean function which represents the circuit shown below (Fig. 10·6).*

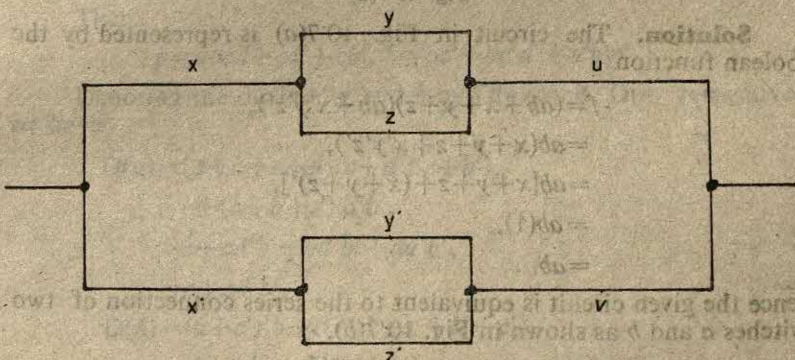


Fig. 10·6

Solution. Observe that :

1. Switches y and z have been connected in parallel. They realize the function $y+z$.

2. Switches/circuits designated by x, $y+z$, and u have been connected in series. They realize the function $x(y+z)u$.

3. Switches y' and z' have been connected in parallel. They realize the function $y'+z'$.

4. Switches/circuits designated by x', $y'+z'$ and v have been connected in series. They realize the function $x'(y'+z')v$.

5. The circuits realizing the functions $x(y+z)u$ and $x'(y'+z')v$ have been connected in parallsl. They realize the function

$$x(y+z)u+x'(y'+z')v$$

10·9·5. Simplification of Circuits

Having seen as to how we can find a function representing a given circuit, and conversely as to how we can draw a circuit which realizes a given function we now take up simplification of circuits using Boolean algebra. The steps involved have already been outlined in section 10·9·4.

Example 16. *Simplify the circuit in Fig. 10·7(a).*

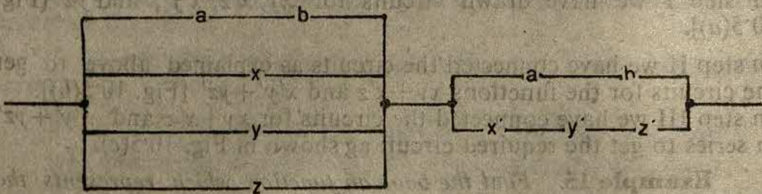


Fig. 10·7(a)

Solution. The circuit in Fig. 10·7(a) is represented by the Boolean function

$$\begin{aligned} f &= (ab+x+y+z)(ab+x'y'z'), \\ &= ab(x+y+z+x'y'z'), \\ &= ab[x+y+z+(x+y+z)'], \\ &= ab(1), \\ &= ab. \end{aligned}$$

Hence the given circuit is equivalent to the series connection of two switches a and b as shown in Fig. 10·7(b).

Fig. 10·7(b)

In applying the usual laws of Boolean algebra to simplify a boolean function f it sometimes happens that a possible simplification is not obvious and therefore it does not occur to us. In such a situation it is useful to take the dual of f and simpilfy it, and then again take the dual to obtain a simplified expression for the original function. The following example will illustrate the method.

Example 17. *Simplify the circuit in Fig. 10·8.*

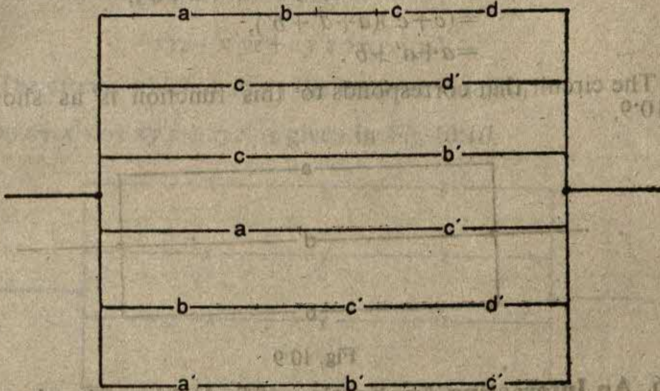


Fig. 10·8

Solution. The circuit is represented by the function

$$f=abcd+cd'+cb'+ac'+bc'd'+a\,b'c'.$$

Consider the first three terms as the function g and the last three terms as the function h.

Then

$$g=abcd+cd'+cb',\ h=ac'+bc'd'+a'b'c' \quad \ldots(1)$$

Denoting the duals of g and h by $D(g)$ and $D(h)$ respectively, we have

$$\begin{aligned} D(g)&=(a+b+c+d)(c+d')(c+b'),\\ &=c+(a+b+d)\,d'b',\\ &=c+ad'b'+bd'b'+dd'b',\\ &=c+ad'b'. \end{aligned} \quad \ldots(2)$$

$$\begin{aligned} D(h)&=(a+c')(b+c'+d')(a'+b'+c'),\\ &=c'+a\,(b+d')(a'+b'),\\ &=c'+(b+d')(aa'+ab'),\\ &=c'+(b+d')(ab'),\\ &=c'+b\,(ab')+d'\,(ab'),\\ &=c'+d'ab'. \end{aligned} \quad \ldots(3)$$

Taking the duals of $D(g)$ and $D(h)$ we have

$$D\,(D(g))=g=c\,(a+d'+b'), \quad \ldots(4)$$

$$D\,(D(h))=h=c'\,(d'+a+b'), \quad \ldots(5)$$

From (1), (4), and (5), we have

$$\begin{aligned} f=g+h&=c\,(a+d'+b')+c'\,(d'+a+b'),\\ &=(c+c')(a+d'+b'),\\ &=a+d'+b'. \end{aligned}$$

The circuit that corresponds to this function is as shown in Fig. 10·9.

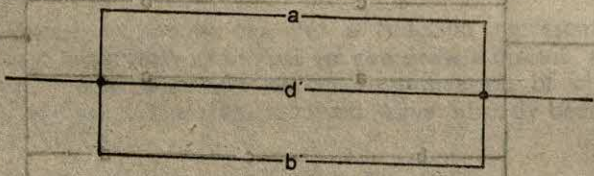


Fig. 10·9

10·9·6. An Interesting Application of Switching Circuits

Example 18. *The Executive Committee of a school consists of three members—the president, secretary, and treasurer, each of whom has control over one of the three switches x, y, z. A resolution is put to vote and is to be decided by a simple majority. Each member votes on the resolution by closing his switch if he votes in favour of the resolution and opening his switch if he votes against the resolution. Design a circuit which will be in state one if and only if the resolution is passed.*

Solution. Let us designate the three members as x, y, z. The resolution will be passed if and only if either all the three members vote for it or exactly two members vote for it. Furthermore, if two members vote for it, they could be either X and Y, or, Y and Z, or, Z and X.

If we denote a vote in favour of the resolution by 1, and a vote against the resolution by 0, then the resolution will be passed if and only if exactly one of the four mutually exclusive sets of conditions is satisfied :

(*i*) $x=y=z=1$.

(*ii*) $x=0, y=z=1$.

(*iii*) $y=0, x=z=1$.

(*iv*) $z=0, x=y=1$.

The above conditions can be rewritten as

(*i*) $xyz=1$,

(*ii*) $x'yz=1$,

(*iii*) $xy'z=1$,

(*iv*) $xyz'=1$,

respectively.

A necessary and sufficient condition for the resolution to be passed is that any one of these conditions is satisfied. Thus the resolution will be passed if and only if

$$xyz+x'yz+xy'z+xyz'=1.$$

The circuit which realizes the boolean function

$xyz+x'yz+xy'z+xyz'$ is given in Fig 10·10.

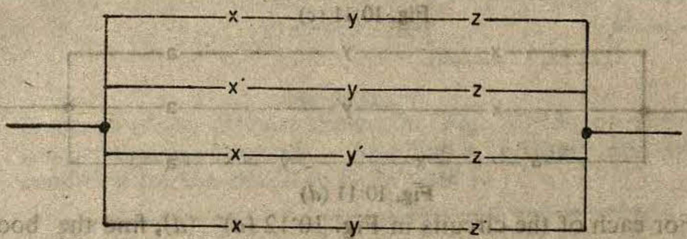


Fig. 10·10

EXERCISE 10 (*g*)

1. Draw circuits which realize each of the following expressions, without first simplifying the expressions :

(*a*) $xy+x'z+yz$ (*b*) $xyz+x'yz+xyz'$

(*c*) $(xy+xz)(x'y+xz')$ (*d*) $x+xyz+x'y$

(*e*) $(x+y)(y+z)(x+z)$ (*f*) $(x+y'z')(y+x'z)(z+x'y')$.

2. For each of the circuits in Fig. 10·11 (*a*)–(*d*), find the boolean function that represents the circuit.

Fig. 10·11 (*a*)

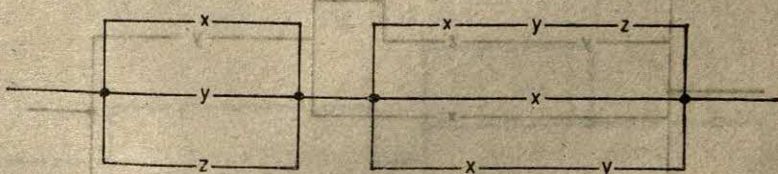


Fig. 10·11 (*b*)

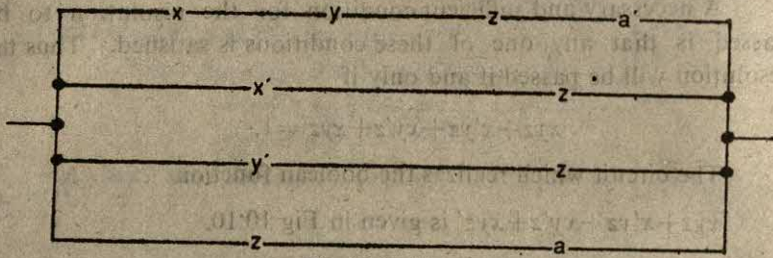


Fig. 10·11 (*c*)

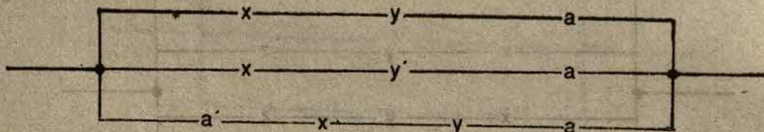


Fig. 10·11 (*d*)

3. For each of the circuits in Fig. 10·12 (*a*)—(*d*), find the boolean function that represents the circuit and simplify it.

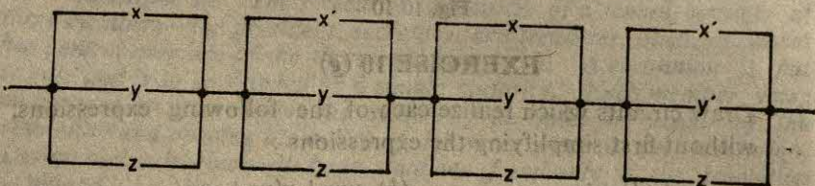


Fig. 10·12 (*a*)

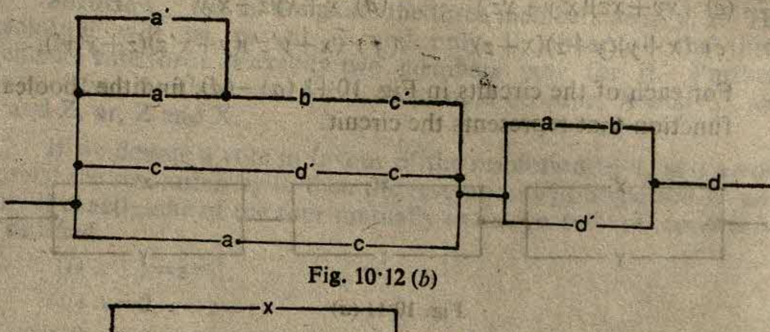


Fig. 10·12 (*b*)

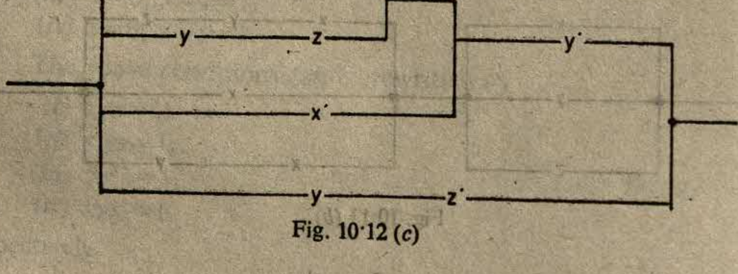


Fig. 10·12 (*c*)

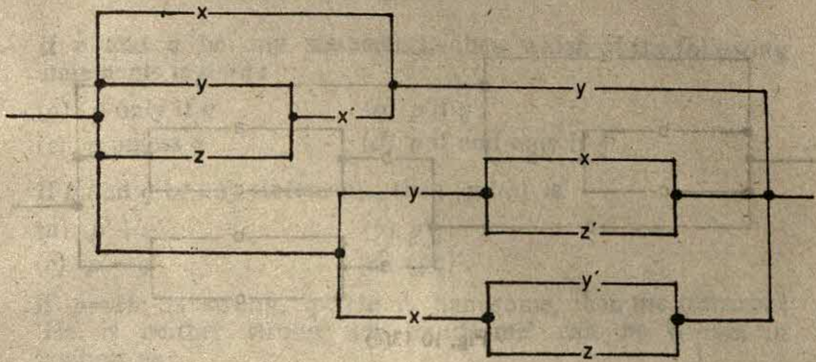


Fig. 10·12 (*d*)

4. For each of the circuits shown in Fig. 10·13 (*a*)—(*e*) find a simpler equivalent circuit. Also find a necessary and sufficient condition for the circuit to be in state 1.

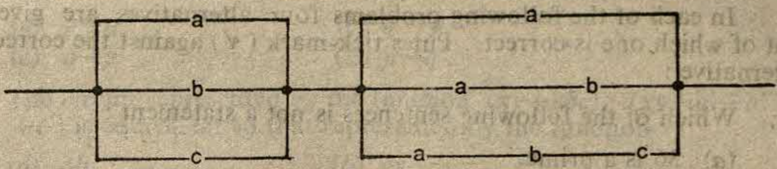
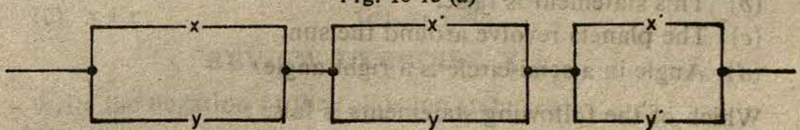


Fig. 10·13 (*a*)



Fig, 10·13 (*b*)

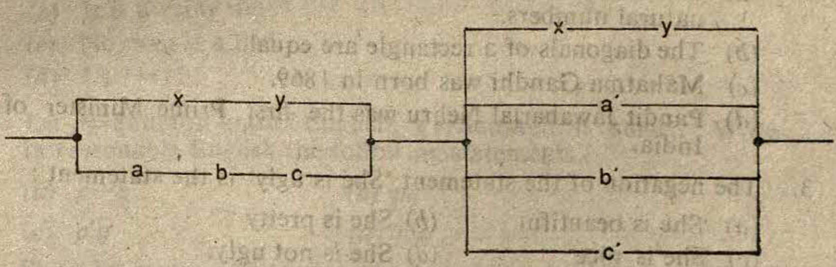


Fig. 10·13 (*c*)

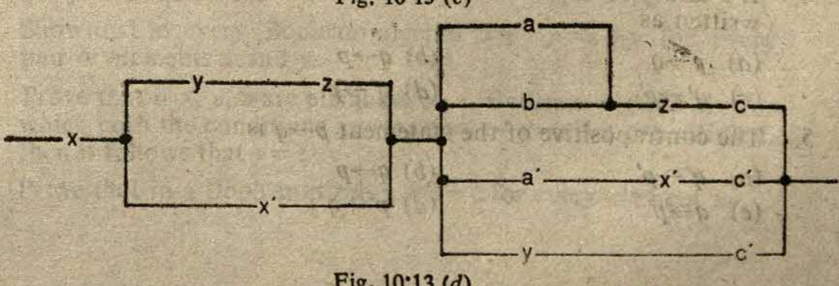


Fig. 10·13 (*d*)

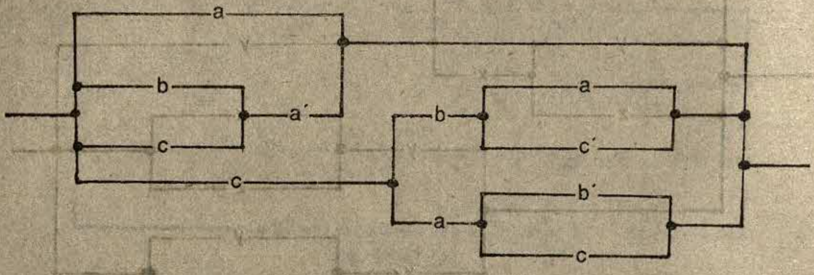


Fig. 10·13(*e*)

5. Verify by drawing circuits, that all the laws of Boolean algebra hold for switching circuits.

TEST YOUR UNDERSTANDING X

In each of the following problems four alternatives are given out of which one is correct. Put a tick-mark (√) against the correct alternative :

1. Which of the following sentences is not a statement :
   (*a*) 36 is a prime.
   (*b*) This statement is false.
   (*c*) The planets revolve around the sun.
   (*d*) Angle in a semi-circle is a right angle.
2. Which of the following statements is false :
   (*a*) 5 cannot be expressed as the sum of squares of two natural numbers.
   (*b*) The diagonals of a rectangle are equal.
   (*c*) Mahatma Gandhi was born in 1869.
   (*d*) Pandit Jawaharlal Nehru was the first Prime Minister of India.
3. The negation of the statement 'She is ugly' is the statement :
   (*a*) She is beautiful (*b*) She is pretty
   (*c*) She is nice (*d*) She is not ugly.
4. If p and q be any statements, then the statement 'p unless q' is written as
   (*a*) $p \to q$ (*b*) $q \to p$
   (*c*) $q' \to p'$ (*d*) $q' \to p$.
5. The contrapositive of the statement $p \to q$ is
   (*a*) $q' \to p'$ (*b*) $q \to p$
   (*c*) $q \to p'$ (*d*) $p' + q'$.

6. If p and q be any statements, then which of the following statements is $p \to q$:

   (a) p only if q (b) p if q
   (c) p unless q (d) p if and only if q.

7. If p and q be any statements, then $(p+q)'$ is

   (a) $p'+q'$ (b) $p'q'$
   (c) $p' \to q'$ (d) $(pq)'$.

8. If p=He is strong, q=He is handsome, then the statement 'He is neither strong nor handsome' can be written in symbols as :

   (a) $p'+q'$ (b) $p'+q$
   (c) $p+q'$ (d) $p'q'$.

9. The converse of the statement $p' \to q'$ is

   (a) $p \to q$ (b) $q' \to p$
   (c) $q \to p$ (d) $p \to q'$.

10. The circuit represented by the function $(xy+abc). (xy+a'+b'+c')$ is equivalent to that represented by the function

    (a) ab (b) xy
    (c) $x+y$ (d) $xy+ab$.

REVIEW EXERCISE X

1. Write the negation of the following statements :

   (a) All roses are red.
   (b) It is a rainy day.
   (c) No man is a liar.
   (d) $1+1=11$.

2. Let p=Saurabh knows English, q=Saurabh is smart. Write in reasonable English the following statements :

   (a) $p+q$ (b) pq
   (c) $p'q$ (d) $p'q'$.

3. Show by means of a truth table that the statements $p \to q$ and $q' \to p'$ are equivalent.

4. Show that in every Boolean algebra $x+x'y=x+y$ for every pair of elements x and y.

5. Prove that if x, y, z are elements of a Boolean algebra B for which both the conditions $xy=xz$ and $x+y=x+z$ are satisfied, then it follows that $y=z$.

6. Prove that in a Boolean algebra $x\,0=0$ for every element x.

7. Show that in a Boolean algebra, for all elements x, y, z, $x'yz+xyz'+x'y'z+x'yz'+x'y'z'=(xy'+xz)'+x'$.

8. Show by writing truth tables that

(a) $[(p\to q)+(p\to r)] \leftrightarrow [p\to(q+r)]$.

(b) $[(p\to q)(p\to r)] \leftrightarrow [p\to(qr)]$.

(c) $[(p\to r)+(q\to r)] \leftrightarrow [(pq)\to r]$.

(d) $[(p\to r)(q\to r)] \leftrightarrow [(p+q)\to r]$.

9. Show that the following arguments are valid :

(a)
$$\begin{array}{l} p\to q \\ r\to q' \\ \underline{r'\to s} \\ \underline{p\to s} \end{array}$$

(b)
$$\begin{array}{l} q\to p \\ q+s \\ \underline{s'} \\ \underline{p} \end{array}$$

10. Draw circuits which realize the following functions :

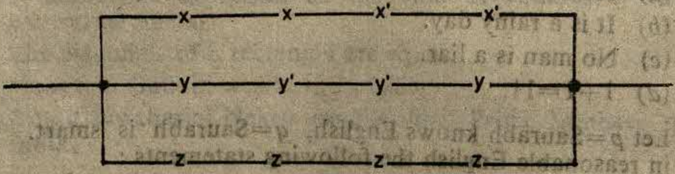
(a) $z(x+y)+x'y'z'$;

(b) $a(x+yz)+d(x'+y)$;

(c) $(xyz+x'yz)(x+yz)$;

(d) $z(uvw+xu'w+y'vw)$.

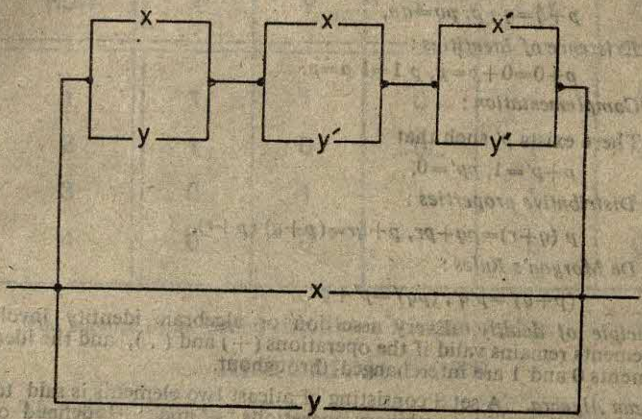
11. Simplify the following circuits :



12. Simplify the following circuits :

(a)

(b)



SUMMARY

1. A sentence of which it is meaningful to say whether it is true or false is called a *statement* A statement must be either true or false but not both.
2. The *negation* of the statement p is the statement 'It is false that p'.' It is denoted by p' (or $\sim p$).
3. The *disjunction* of p and q is 'p or q'. It is denoted by $p+q$ (or $p \vee q$).
4. The *conjunction* of p and q is 'p and q'. It is denoted by pq (or $p \wedge q$).
5. A statement which is always true is called a *tautology*. It is denoted by 1.
6. A statement which is always false is denoted by 0.
7. The statement 'If p, then q' is called the *implication* or a *conditional statement*'. It is defined as $p'+q$, and is denoted by $p \rightarrow q$.
8. The statement 'p if and only if q' is called an *equivalence* or a *biconditional statement*. It is defined as $(p \rightarrow q)(q \rightarrow p)$.
9. Truth table for p', $p+q$, pq, $p \rightarrow q$, $p \leftrightarrow q$.

Row	p	q	p'	p + q	pq	p → q	q ↔ p
1	1	1	0	1	1	1	1
2	1	0	0	1	0	0	0
3	0	1	1	1	0	1	0
4	0	0	1	0	0	1	1

10. If p, q, r be any statements, then the following properties hold :

(a) *Commutative properties* :

$$p+q=q+p,\ pq=qp,$$

(b) *Existence of identities* :

$$p+0=0+p=p,\ p\,1=1\,p=p.$$

(c) *Complementation* :

There exists p' such that

$$p+p'=1,\ pp'=0.$$

(d) *Distributive properties* :

$$p\,(q+r)=pq+pr,\ p+qr=(p+q)\,(p+r).$$

(e) *De Morgan's Rules* :

$$(p+q)'=p'q',\ (pq)'=p'+q'.$$

11. *Principle of duality.* Every assertion or algebraic identity involving statements remains valid if the operations (+) and (.), and the identity elements 0 and 1 are interchanged throughout.

12. *Boolean Algebra.* A set S consisting of atleast two elements is said to be a Boolean algebra for two binary operations '+' and ' . ' defined on it if the following properties B1—B5 are satisfied :

B 1. *Closure.* S is closed for the operations '+' and ' . '.

B 2. *Commutativity.* The operations '+' and ' . ' are commutative.

B 3. *Identities.* There exist in S distinct identity elements 0 and 1 relative to the operations '+' and ' . ' respectively.

B 4. *Distributivity.* For each x is S, there exists an element x' in S such that

$$x+x'=1,\ xx'=0.$$

The following properties also hold in a Boolean algebra :

B 6-B 12. For all elements x, y, z in a Boolean algebra

$$x+x=x,\ xx=x.$$

$$x+1=1,\ x\,.\,0=0.$$

$$x+xy=x,\ x(x+y)=x.$$

$$x+(y+z)=(x+y)+z,\ x\,(yz)=(xy)\,z.$$

$$(x+y)'=x'y',\ (xy)'=x'+y'.$$

$$0'=1,\ 1'=0.$$

$$(x')'=x.$$

13. *Rule of detachment* (modus ponens). The propositions 'p' and '$p\to q$' together yield the proposition q.

14. *Law of syllogism.* The propositions '$p\to q$' and '$q\to r$' together yield the proposition '$p\to r$'.

15. A switch is said to be *closed* if current can pass through it. It is said to be *open* if current cannot pass through it.

A closed switch is said to be in state 1 and an open switch is said to be in state 0.

16. If two switches are connected in *parallel*, the resulting circuit is denoted by $x+y$. If two switches are connected in *series*, the resulting circuit is denoted by xy.

17. Closure properties of x', $x+y$, xy are given by the following table :

Row	x	y	x'	x +y	xy
1	1	1	0	1	1
2	1	0	0	1	0
3	0	1	1	1	0
4	0	0	1	0	0

18. The algebra of switching circuits is a Boolean algebra.

HISTORICAL NOTE

The development of mathematical logic as an algebra of classes which began with the works of George Boole (1815-1864)—*The Mathematical Analysis of Logic* (1847) and *An Investigation of the Laws of Thought* (1854) in the nineteenth century. It was followed by the pioneering work of Gotlieb Frege. Frege's work went unnoticed until Bertrand Russel drew attention to it during the first decade of the present century. The method of truth tables owes its origin to C.S. Pierce in the 1870's but was developed fully later on by E. Post and Ludwig Wittgenstein. The principle of duality was discovered independently by De Morgan and Benjamin Pierce.

BLAISE PASCAL (1623-1662)

French mathematician, Pascal, was a mathematical prodigy. At the age of 16, he published a one-page paper, which page is a most fruitful page in the history of mathematics. The result contained in that paper is now known as Pascal's theorem (viz-points of intersection of opposite sides of a hexagon inscribed in a conic are collinear). Pascal's second love (the first being geometry) was designing, building and selling calculating machines after which he turned to hydrostatics in 1648. He made himself a name in this field through the well-known Puy-de-Dome experiment confirming the weight of air. About six years later, Pascal again turned to mathematics. At this time of his life, he wrote '*Complete Work on Conics*' which unfortunately is not extant, and it was at this stage that Pascal became interested in the theory of probability. Using arithmetic triangle (now known after Pascal's name), he extended Cardan's results further. Soon afterwards, Pascal turned to theology which ended his mathematical career except for one later flash, when on one jolly night in the year 1658, he had a toothache and to distract himself he started studying the cycloid, obtaining a number of most remarkable results about the same. He published many of these results in his "Treatise on the sines of a Quadrant of a Circle" and according to Leibnitz, it was in this work of Pascal, that he saw the light of Calculus.

CHAPTER 11

Elements of Probability

11·1. INTRODUCTION

Cricket is perhaps the most popular game these days in our country. Some people may doubt this statement. But surely no one can doubt the following statements :

1. It is very unusual for a cricket player to have played in more than 100 Test matches. (By the way, in 1976, M.C. Cowdrey had already played in 114 Tests.)

2. It is very unlikely for a sparrow to die of a hurt caused by a ball during a Test match. (By the way, a sparrow was killed in flight by a ball from M. Jahangir Khan at Lord's in 1936.)

3. The probability of a team making more than 4 centuries in a single innings of a Test match is very small. (By the way, Australia made 5 centuries in one innings at Kingston in June 1955 while playing against West Indies.)

4. The chances of our country winning a Test match played at Bombay are very bright because most of the times our team has been doing well there.

We believe in the above statements because of our past experience with the frequency of the occurrence of the events mentioned. The rarer the event, the lesser our faith in its repetition.

Let us consider another question related with Test matches. How do they decide which team is going to play first ? They toss a coin to settle the question and no objection is raised to this method by either team. The reason is obvious. It is a fair method in the sense that neither of the teams has more advantage or disadvantage as compared with the other ; *their chances of winning the toss are exactly the same*. You know this fact by insight which capitalizes on the *symmetry* of the situation or the fact that both the head or the tail are *equally likely* without ever having had any instruction in the so-called mathematics of chance or the *theory of probability.*

In order to know the meaning of the word *Calculus,* you had to look up the dictionary (if at all you bothered !), but the word *probability* must be as familiar to you as *line* and *point* were, before you had ever studied Geometry. To convince you further about the

fact that the rudiments of the probability theory are already dormant in you, let us play a simple game.

Let us take a jar and put some marbles into it. You will take out a marble from the jar without looking into it. If this marble turns out to be green, you win ; if it is red, we win. Let us start with 8 red marbles in the jar. What are your chances of winning ? *Absolutely none* ! O.K. Let us not be so unfair to you. Let us put 2 green marbles also into the jar. What are your changes of winning now ? Very slim ! Let us put in two more green marbles into the jar. There are now 8 red and 4 green marbles in the jar. What are the chances of your winning now ? Better than before. Suppose we put 4 more green marbles into the jar so that there are 8 red and 8 green marbles in the jar. What are your chances of winning now ? *Fifty-fifty* ! If I now remove all the 8 red marbles from the jar leaving the green ones in and then let you take out a marble, what would be your chances of winning ? You are sure to win, eh ! You have a *cent-per-cent* chance of winning ; do you not ? Except at the beginning of the game (when you were sure to lose) and at the end (when you were sure to win), there was an element of uncertainty as to the outcome of the game. At every stage in between, either a red or a green marble could bave been drawn ; even when there were 8 red and 2 green marbles, a green marble could have been the result of our draw ; yet you were prepared to say that the chances of your winning were very slim.

The situation under consideration is hypothetical but surely we make so many predictions, take decisions and plan accordingly regarding so many situations in every sphere of life everyday, the outcomes of which are quite *uncertain*. For example, on a cloudy day during the rainy season, we pick up our umbrella/raincoat while going out even when we are not certain that it is going to rain. So many of us purchase a lottery ticket even when we are not sure of a prize. The weather-forecasters predict temperatures, rainfall, humidity, storms etc. Number of likely accidents ; frequency of likely crimes ; predictions about births and deaths, marriages and divorces ; figures regarding employment, stocks, sales, income-tax, interest rates etc during an on-coming period ; likely results of any type of election are all examples of our dealing with situations whose outcomes are uncertain. Yet we predict, decide, plan and go into action. The science that helps us here is the *theory of probability*.

The *theory of probability* is the science that enables us to deal with uncertainty in an effective way. It is a mathematical scicence in as much as it creates its own models of uncertainty, and works within its own laws. Of course, to be useful, the models are created to approximate essential features of real life situations. The results of theory are then applied to actual problems with satisfactory results.

Going back to our game of marbles, we observe that we used such phrases as very slim, better, fifty-fifty chance, of a cent-per-

cent chance of winning. Thus it appears that the chance or probability of our winning is something which admits of comparison in magnitude. Naively, we assigned numerical values also for the purpose of comparison to our chance or the probability of our winning. We shall now try to build up the tools which will help us in the difficult task of assigning numerical values to an abstract concept like probability, as scientifically as we may manage.

11·1·1. Random Experiments and Sample Space

Here is an interesting experiment. Take a tumbler. Bring a jug, full of water. Go on puring water from the jug into the tumbler. *Go on* till the jug is empty. No kidding. What happened ? Water spilled all around the tumbler ? But then, did you expect something else ? You knew already that if you kept on pouring water, first the tumbler would get filled, and then the water would start spilling !

Now roll a die. What would show up ? Assuming you are not the *Shakuni* from the old epic *Mahabharat*, you cannot really predict which of the six faces 1, 2, ..., 6 would show up. It might be a *one* or a *two* or any one of *three*, *four*, *five* and *six*. The outcome of rolling the die is not *certain* ; there is some *randomness* associated with it. This experiment is different from the earlier one in that its outcome is not *definite* ; it is *random*. You cannot say what the outcome is going to be. However, you cay say that it is going to be one of the six outcomes mentioned above.

Suppose that each letter of the word HONEY is written on a card and the cards are shuffled. A card is then drawn at random, the letter on the card is noted, the card is returned into the deck and the deck is shuffled again. We may keep on drawing one card in this fashion and go on noting the letter written on it as many times as we please. Each time we draw a card, we do not know which letter it is going to bear, but we do know that *it must be one and only one element of the set {H, O, N, E, Y} every time.*

Let us consider another situation. Suppose we take a fifty paisa coin. It has two faces, one showing the national emblem and the other showing its denomination. Let us agree to call the former one head (H) and the latter one tail (T). If we toss, either 'H' or 'T' would show up. These two are the possible outcomes, one would normally imagine, of our coin throwing. We may list them in a set as {H, T}. On course, the *hair-splitter* may argue that the coin might stand on its edge (E), (As a matter of fact, if a coin is specially made so that its thickness equals one-third of its diameter, then it will have almost as much chance of standing on an edge as it has of showing a head or a tail) or roll away (R), or be lost (L) by being tossed out of sight. Allowing for these fuss-makers, we may list the set of all possible or conceivable outcomes of our throw as {H, T, E, R, L}. We may repeat out action of throwing the coin under the same conditions

as above as many times as we like ; *the outcome is going to be always one and only one element of the set we have listed above, though we do not know which one.*

Drawing of the card or tossing of the coin an explained above, are examples of what is known as a *random experiment.* The essential features of such an experiment are :

1. We agree upon the conditions under which the experiment is to be performed and that it can be repeated under similar or fairly uniform conditions.

2. We agree upon the possible outcomes of the experiment. Each time we perform the experiment, we do not know what the outcome might be (*it depends upon chance* !) but it must be one of the outcomes agreed upon.

Throughout this chapter and the next, we shall use the term *experiment* to mean *random experiment* unless otherwise stated.

Definition. *For any experiment E, let S be a set of outcomes, such that*

(*i*) *each element of S denotes a possible outcome of the experiment, and*

(*ii*) *whenever the experiment is repeated, the outcome is one and only one element of S.*

Then S is called a **sample space** *of the experiment.*

Every element of S is called a **sample point**.

Notation. We shall denote the sample space of an experiment by

$$S=\{S_1, S_2, \ldots\ldots\ldots\ldots S_n\},$$

where $S_1, S_2, \ldots\ldots\ldots S_n$ are the various sample points.

A sample space is called **finite** or **infinite** according as the number of sample points is finite or not. A sample space is called **discrete** if either it is finite or else it is countably infinite (*i.e.*, it is infinite but its elements are in one-to-one correspondence with the set of natural numbers). In this text, we shall be dealing with finite sample spaces only.

Remarks 1. We have used the preposition '*a*' before the word sample space because it is possible to have more than one sample space for the same experiment. In our coin-tossing experiment, we have listed two distinct sample spaces {H, T} and {H, T, E, R, L}. In a given situation, one sample space may be preferable to another on the grounds of usefulness, computational ease or even intuitive appeal.

2. The words space and point here, are not used in their physical or geometrical sense. It is an altogether different matter

that just as we represent sets geometrically by Venn diagrams, we may have graphic representation of sample space too ! Remember, a sample space is after all a *set.*

3. A sample space represents the experiment into the possible outcomes. Thus, once a sample space is agreed upon, we can forget about the experiment, so far as our further work is concerned.

4. Impossible outcomes may be included in a sample space. The point to remember is that all *possible ones* must be included. See Experiment 7 further on.

5. No set rules for selecting the sample points can be laid down. Sample points are the undefined objects of the theory.

Here are some examples :

Experiment 1. A card drawn from a pack of cards, for its suite (heart '*h*', diamond '*d*', club '*c*', or spade '*s*').

Sample space : $S=\{h, d, c, s\}$.

Experiment 2. Drawing a ball, for its colour, from an urn containing three balls of different colours—red (r), blue (b) or green (g).

Sample space : $S=\{r, b, g\}$.

Experiment 3. Tossing a coin twice, for number of times an 'H' turns up.

Sample space : $S=\{0, 1, 2\}$.

The points of the sample space may conveniently be represented by the points (0, 0), (1,0) and (2, 0) on the X-axis.

Experiment 4. From an urn containing 3 red balls, 2 blue balls and 2 green balls, drawing two balls in succession, for colour.

Sample space :

$S=\{rr, rb, rg, br, bb, bg, gr, gb, gg\}$, where rb, for example, means that the first ball drawn was red and the second one blue. Similarly for other symbols.

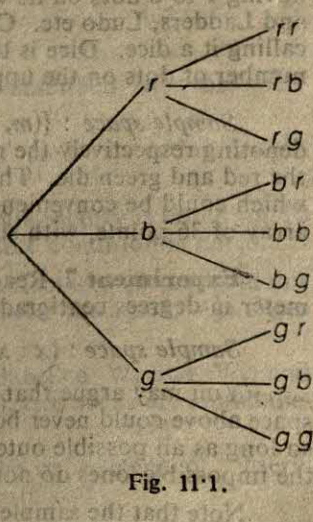


Fig. 11·1.

Remarks 1. In a two (or more than two) step experiment as above, a tree diagram as shown below, is a great help in imagining all the possible outcomes.

2. Another sample space, and a more useful one at that, concerned with this experiment is obtained as follows. Label the

three red balls R_1, R_2, R_3; the two blue balls B_1, B_2, and the two green ones G_1, G_2. Suppose that the first ball drawn is R_2 and that the second ball drawn is G_1; this outcome, let us denote it by (R_2, G_1). Similarly for other possible outcomes. The sample space is :

$$
\begin{array}{llllll}
\{(R_1, R_2), & (R_1, R_3), & (R_1, B_1), & (R_1, B_2), & (R_1, G_1), & (R_1, G_2), \\
(R_2, R_1), & (R_2, R_3), & (R_2, B_1), & (R_2, B_2), & (R_2, G_1), & (R_2, G_2), \\
(R_3, R_1), & (R_3, R_2), & (R_3, R_1), & (R_3, B_2), & (R_3, G_1), & (R_3, G_2), \\
(B_1, R_1), & (B_1, R_2), & (B_1, R_3), & (B_1, B_2), & (B_1, G_1), & (B_1, G_2), \\
(B_2, R_1), & (B_2, R_2), & (B_2, R_3), & (B_2, B_1), & (B_2, G_1), & (B_2, G_2), \\
(G_1, R_1), & (G_1, R_2), & (G_1, R_3), & (G_1, B_1), & (G_1, B_2), & (G_1, G_2), \\
(G_2, R_1), & (G_2, R_2), & (G_2, R_3), & (G_2, B_1), & (G_2, B_2), & (G_2, G_1)\}.
\end{array}
$$

The elements of this sample space are equally likely, whereas those of the earlier one are not. (See page 627 for equally likely outcomes.)

Experiment 5. From the urn of Experiment 4, drawing two balls simultaneously, for colour.

Sample space. $S=\{rr, rb, rg, bb, bg, gg\}$.

Experiment 6. Rolling a red die (Like the cubical piece having 1 to 6 dots on its various faces, with which you play Snakes and Ladders, Ludo etc. Correct yourself if you are in the habit of calling it a dice. Dice is the plural of die.) and a green die for the number of dots on the upper-most faces.

Sample space : $\{(m, n) : 1 \leqslant m \leqslant 6,\ 1 \leqslant n \leqslant 6\}$, m and n denoting respectively the number of dots on the uppermost face of the red and green die. The sample space consists of 36 elements which could be conveniently represented geometrically by a double array of 36 points, with six rows and six columns.

Experiment 7. Reading daily temperature on a room thermometer in degrees centigrade.

Sample space : $\{x : x$ is a positive real number$\}$.

You may argue that many of the values listed in the sample space above could never be obtained. This should not worry us; so long as all possible outcomes are included in the sample space, the impossible ones do not matter.

Note that the sample space is infinite here.

Experiment 8. Noting the number of times a coin has to be tossed before a head is obtained.

Sample space. $\{0, 1, 2, 3, \ldots\ldots\ldots, n, \ldots\ldots\ldots\ldots\}$.

11·1·2. Events

Our interest in the sample space lies mainly due to the fact that it enables us to describe events which can occur as a result of our experiment. Thus as a result of our coin-tossing experiment, the event of a head turning up may occur. It corresponds to the sample point H or the subset {H} of the sample space {H, T}. As a result of a die-rolling experiment, the event of there being an even number of dots on the uppermost face may occur. It corresponds to the sample points 2, 4 and 6 or the subset {2, 4, 6} of the sample space {1, 2, 3, 4, 5, 6}. We define an event as follows :

Definition. *Any subset of the sample space is called an event.*

In particular, the sample space S, being a subset of itself, is also an event. ϕ is for the same reason, an event. They are known respectively as the **sure** and the **impossible** events.

The singleton $\{S_i\}$, where $S_i \in S$, is also an event, for each i. These events corresponding to the lone sample points are called **elementary** or **simple** events. In contrast to elementary events, events other than elementary, are called **composite** events, but we shall hardly ever add the adjective.

Example 1. In Experiment 3, the event, that the head turns up at least once, is {1, 2}.

Example 2. In Experiment 4, the event that both the balls have the same colour, is $\{rr, bb, gg\}$.

Example 3. In Experiment 6, the event that the sum of the number of dots on the two dice is less than four is, {(1, 1), (1, 2), (2, 1)}.

EXERCISE 11 (*a*)

Write a sample space for experiments described in problems 1 to 4 below :

1. 500 seeds are sown and the number of seeds sprouting within 7 days is noted.
2. A sewing machine produced in a certain factory is tested for 10 parts and the number of parts in perfect condition is noted.
3. It is known that a batch of 100 manufactured items contains 10 defective items. Two items, one after the other, are taken out and tested for being defective.
4. From a box containing two white (labelled W_1 and W_2) and two pink socks (labelled P_1 and P_2), two socks are simultaneously drawn.
5. What sort of a sample space would the experiment of observing the length of life of electron tubes have ?
6. List all the elementary events for the experiment in Exercise 2 above.

7. Three six-faced cubical dice are rolled and the sum of the numbers shown up is added. Write a sample space for this experiment. What is the event that this sum
   (*a*) exceeds 12 ? (*b*) is less than 6 ?
8. What is the event that in the experiment of Exercise 1 above, the number of seeds sprouting is
   (*a*) less than 100 ? (*b*) more than 50 ?
   (*c*) not less than 10 and not more than 490 ?
   (*d*) zero ? (Be careful ! The answer is not 0 !)
   (*e*) less than 550 ? (*f*) more than 600 ?
9. Can the events in 8(*a*) and 8(*b*) occur simultaneously ? Can those in 8(*b*) and 8(*d*) occur simultaneously ?
10. Let S be the sample space of outcomes when an ordinary cubical die is thrown once. Let A and B be respectively the events of an even and an odd number of dots on the uppermost face. List A and B. Most of the events A and B occur ?

11·2. ALGEBRA OF EVENTS

Since, used in a probabilistic context, we have called a subset of a sample space an event, therefore, just like the algebra of subsets of a set, we can develop the algebra of events that may happen as the result of a random experiment. Recall that you talked about equality, complementation, union and intersection in the context of sets. We shall do the same in case of events.

Suppose S is a sample space and E_1, E_2 are two events. E_1 are E_2 are said to be **identical** or **equivalent** (denoted as $E_1=E_2$), provided, E_1 occurs if and only if E_2 occurs.

Example 4. In drawing two cards from an ordinary pack of playing cards, let E_1 be the event that *both the cards are picture cards* and let E_2 be the event that *the face values of cards add up to at least 22*. If the face value of the ace is taken as 1 and it is not to be regarded as a picture card, then $E_1=E_2$.

Consider the random experiment of rolling a six-faced cubical die maked with one to six dots on various faces. The sample space may be taken as $S=\{1, 2, 3, 4, 5, 6\}$. Let E be the event *an even number of dots on the uppermost face* when the die is rolled. Then $E=\{2, 4, 6\}$. Corresponding to this event E, there is the event *NOT an even number of dots on the upper-most face*. Clearly this may be called the event "not E". It is represented by the subset $\{1, 3, 5\}$ of S. Note that $\{1, 3, 5\}$ is the set-complement of E in S. This is not surprising because "not E" must contain those and only those points of the sample space which do not correspond to E. The event "not E" is accordingly denoted by $\sim$ E, the wiggle ($\sim$) denoting complementation, and is also known as the event *complementary to* or *opposite of* E.

Other symbols used to denote "not E" are E′, $\overline{E}$ and S ~ E, S being the sample space. Thus we have the following :

Definition. *If E is a given event on the sample space S, then the complementary or opposite event ' NOT E" is the event given by the complement of E in S.*

Since S ~ (S ~ A) = A for all subsets A of S, it is clear that if B = "Not A", then A = "Not B". Hence the relation of being opposite of an event is symmetric.

Thus instead of saying *A is the event opposite of B,* we may say *A and B are opppsite events.*

Example 5. In the die-rolling example above, define the events E and F as

E = at most 2 dots, and
F = at least 3 dots.
Then $\bar{E}$ = F and $\bar{F}$ = E, because
E = {1, 2}, F = {3, 4, 5, 6} and so
~E = F, ~ F = E.

Consider now the following events in our die rolling experiment :

A = {2, 4, 6} (*i.e.*, the outcome is even.)
B = {2, 3, 5} (*i.e.*, the outcome is a prime.)
C = {2} (*i.e.*, the outcome is even and also a prime.)
D = {2, 3, 4, 5, 6} {*i e.*, the outcome is either even or a prime.)

You cannot fail to see that the events C and D are related to the events A and B in a certain manner. C warrants the occurring of both A AND B ; as such C is labelled as the event "A AND B" and is nothing but A∩B.

Definition. *If A and B are two events, then their intersection or the event "A AND B" is the event A∩B (also denoted AB).*

Similarly, since for the occurrence of D, the occurrence of A *OR* B is enough, the event D is called the event "A OR B" and is nothing but A∪B.

Definition. *If A and B are two events, then their* **union** *or the event* "A OR B" *is the event A∪B (also denoted A+B).*

Notice that 'or' above is used in the *inclusive* sense ; we may have *either A, or B, or both.* The *OR here does not mean exactly one of A and B ; it means at least one of A and B.*

If A∩B = ϕ, then A and B are known as **mutually exclusive** events. The effect of a sample point being in one is to *exclude* it from the other. Since complementary events have no common sample points, they are mutually exclusive.

If $A \cup B = S$, then A and B are known as **exhaustive** events. Between the two of them, they *exhaust* all the sample points. Complementary events are exhaustive. (Why ?)

Example 6. A card is drawn from an ordinary pack of playing cards. Events E_1 to E_6 are described below :

E_1 : the card drawn is a heart.

E_2 : the card drawn is a diamond.

E_3 : the card drawn is red.

E_4 : the card drawn is black.

E_5 : the card drawn is a queen.

E_6 : the card drawn is the queen of hearts.

Then $E_1 \cap E_5 = E_6$; E_3 and E_4 are mutually exclusive as well as exhaustive ; $E_1 \cup E_2 = E_3$.

EXERCISE 11 (*b*)

1. If A is the event that exactly one of three tested radio-sets is defective and B is the event that all three of them are acceptable (*i.e.*, not defective), what are the events $A \cup B$ and $A \cap B$?

2. One ticket counterfoil is drawn from a bowl for the purpose of giving a lucky prize at a Fair. Suppose 105 tickets (numbered 1 to 105) only were sold. Suppose E is the event that the ticket counterfoil bears a number less than 53 and that F is the event that it bears an even number. Write down a suitable sample space and describe the events $\sim E$, $E \cap F$ and $E \cup F$.

3. Two dice are rolled. A is the event that the sum of the numbers shown by the dice is 3. B is the event that at least one of the dice shows up a 4. Are the events A and B

   (*a*) mutually exclusive ?

   (*b*) exhaustive ?

   Give arguments in support of your answer.

4. A number is chosen from among the numbers 1 to 8, both inclusive. A is the event that a number greater than 3 is chosen and B is the event that a number less than 3 is chosen. Write down a sample space and the events $\sim A$, $\sim B$, $A \cap B$, $A \cup B$, $(\sim A) \cap B$, and $(\sim A) \cup B$. Are any of these events which you have just listed, mutually exclusive or exhaustive ?

5. Show by means of Venn diagrams the events E, $\sim E$, $E \cap F$, and $E \cup F$ in a sample space. In case of $E \cap F$, consider the cases when E and F are mutually exclusive and when they are not. Similarly, draw two diagrams for $E \cup F$, one for the case

when E and F are exhaustive and the other for the case when they are not exhaustive.

6. Show that complementary events are exhaustive but that the converse is not true.

7. If E and F are two events in a sample space S, then show that:

(i) $\sim(E\cup F)=(\sim E)\cap(\sim F)$.

(ii) $\sim(E\cap F)=(\sim E)\cup(\sim F)$.

11·3. PROBABILITY FUNCTION ON A SAMPLE SPACE

Our interest in all possible outcomes of an experiment is not just for the fun of it. As stated earlier, we are interested in knowing which outcome has greater chance of occurring and which one less. This we do by assigning to each outcome S of the sample space, a number P(S). Since our object is to compare only, the numbers P(S) are going to be some ratios and for reasons to be seen soon, we so arrange matters that the sum of the numbers assigned to all the outcomes is unity. We shall confine ourselves to finite discrete sample spaces only in this section.

Let us go back to the discussion of our game of marbles. Let us tabulate that discussion in the form of the table given below. To begin with, do not look at the last column.

Table 11·1.

No. of red marbles in the jar	*m=No. of green marbles in the jar*	*n=total number of marbles in the jar*	*Chances of your winning*	*The ratio m/n*
8	0	8	Absolutely none	0
8	2	10	Very slim	$\frac{1}{5}$
8	4	12	Better	$\frac{1}{3}$
8	8	16	Fifty-fifty	$\frac{1}{2}$
0	8	8	Cent-per-cent	1

Let us now try to assign numerical values to our chance or probability of winning. Surely, you are going to express a "cent-per-cent" chance as 1. (Revise your topic of percentages in Arithmetic if you disagree !) A "fifty-fifty" chance means $\frac{1}{2}$. "Absolutely none" would, no doubt, have a numerical measure zero. Let us now

consider the remaining situations. It is clear that greater the percentage of green balls in the jar, the greater your chance of winning. In fact, your chances of winning increase exactly as this percentage does. "Very slim" may, therefore, be interpreted as 20% or $\frac{1}{5}$ (there are 2 green balls out of 10 at this moment) and "Better" as $33\frac{1}{3}\%$ or $\frac{1}{3}$ (there are 4 green balls out of the total 12). Now look at the last column of the table. The probability of winning is the ratio m/n. In the terminology we have built up, drawing a ball from the jar at every stage may be regarded as an experiment. Your winning is the event that a green ball is drawn. Assigning numerical values to your chance of winning may be regarded as determining the probability of the event that a green ball is drawn. This intuitive approach to determining the probability of an event cannot be adopted universally, but one thing is certain. No event can have a more than cent-per-cent chance of happening and hence we would not want the probability of any event to be greater than 1. It is also clear that the probability of an event cannot be less than zero.

Let us have some more (and simpler yet) illustrations. We shall denote the probability (the chance of happening) of an event E by the symbol P(E). The probability of an elementary event $\{S_i\}$ in a sample space will be denoted by $P(S_i)$ rather than by $P(\{S_i\})$.

Experiment 9. A coin is tossed once, for head or tail.

Sample space : $S=\{H, T\}$.

We believe that the coin is unbiased or fair, *i.e.*, it has as much tendency to turn up 'H' as it has to turn up 'T'. Therefore, we must have

$$P(H)=P(T).$$

As before, the events (H and T have in this case a fifty-fifty chance of occurring. We may, therefore, agree to assigning a value $\frac{1}{2}$ to each of them.

Thus $P(H)=P(T)=\frac{1}{2}$.

Observe that

$$P(H)+P(T)=1.$$

Experiment 10. Suppose we write the letters of the words, CAN on cards and pick up a card at *random* (without any bias towards any one !), and note the letter on it.

Sample space : $S=\{C, A, N\}$.

Obviously, each letter has the *same* chance of being selected. Because of this equal likelihood, we must have

$$P(C)=P(A)=P(N).$$

If Q and R are respectively the events $\{C, A\}$ and $\{C, A, N\}$, our intuition (and reason too !) suggests that

$$P(Q)=2P(C),$$

and $$P(R)=3P(C). \qquad ...(i)$$

Also, the event R is sure to happen so that we may assign a probability 1 to it. But if $P(R)=1$, then in view of (1),

$$P(C)=\tfrac{1}{3}.$$

Hence we must have

$$P(C)=P(A)=P(N)=\tfrac{1}{3}.$$

Note that in this experiment also, the sum of the probabilities of the elementary events is 1.

Another observation you might have made is that in all the above experiments, *each outcome was as likely to occur as any other*; none had more chance of occurring than the rest. We formalize this observation by saying that in each experiment above, all the outcomes were **equally likely**. The observations made so far, initiate the following remark :

Remark. In both the experiments above, you may observe that,

(*i*) the number assigned to a sample point (called the probability of the corresponding elementary event) is always non-negative,

(*ii*) the sum of the numbers assigned is 1,

(*iii*) the numbers are assigned according to some rule, laid down by us, depending upon our knowledge about the outcomes (which was equal-likelihood).

These observations lead us to the following definition of the probability function on a finite discrete space.

Definition. *Let $S=\{S_1, S_2, \ldots\ldots, S_n\}$ be a finite discrete sample space. A probability function on S is a function on S that assigns to each $S_i \in S$, a non-negative real number $P(S_i)$, such that*

(*i*) *for each i,*

$$0 \leqslant P(S_i) \leqslant 1.$$

(*ii*) $P(S_1)+P(S_2)+\ldots\ldots+P(S_n)=1.$

The number $P(S_i)$ assigned to $S_i \in S$, is also called the probability of the elementary event $\{S_i\}$, $i=1, 2, \ldots\ldots, n$.

If all the outcomes are equally likely, then we shall call the probability function the **equiprobable** (or **natural**) **probability function**.

As in Experiments 9 and 10 above, it may happen in numerous other cases that each outcome is as likely to occur as any other. It, therefore, looks fair (natural) in such cases to assign equal probabilities to them. Thus in a sample space with equally likely outcomes, we must have

$$P(S_i)=P(S_j),\ S_i,\ S_j \in S, \text{ for all } i, j.$$

Theorem. *Let $S=\{S_1, S_2, \ldots\ldots, S_n\}$ be a sample space. If P is the equiprobable (natural) probability function on S, then*

$$P(S_i)=\frac{1}{n} \text{ for all } i=1, 2, \ldots\ldots, n.$$

Proof. Since P is equiprobable on S,

$$P(S_i)=P(S_j) \text{ for all } i, j=1, 2, \ldots\ldots, n.$$

But

$$P(S_1)+P(S_2)+\ldots\ldots P(S_n)=1.$$

$$\therefore \quad P(S_i)=\frac{1}{n},\ i=1, 2, \ldots\ldots, n.$$

11·3·1. Probability of an Event. Just as we are interested in knowing the likelihood (greater or less) of an outcome to occur, as compared to that of others, so are we in the comparison of the likelihood of events. However, having assigned the probabilities to the elementary events, the rules of the game, so to say, have already been set up, and the matter of assigning probabilities to arbitrary events is straightforward.

Definition. *Let $S=\{S_1, S_2, \ldots\ldots, S_n\}$ be a sample space, and P a probability function on S. Let $A \subset S$, be an event. We define the probability of the event A, to be denoted by $P(A)$, as follows:*

(*i*) *If $A=\phi$, then $P(A)=0$.*

(*ii*) *If $A \neq \phi$, and $A=\{S_1, S_2, \ldots\ldots, S_k\}$, then*

$$P(A)=P(S_1)+P(S_2)+\ldots\ldots P(S_k).$$

It follows in particular, that

$$P(S)=1.$$

11·3·2. Probability of an Event on the Assumption of Equilikely (or Equiprobable) Outcomes

Theorem. *If $S=\{S_1, S_2, \ldots\ldots, S_n\}$ be the sample space, the outcomes being equally likely, then for any event $A \subset S$, having m elements, $P(A)=\dfrac{m}{n}$.*

Proof. Since the outcomes are equally likely, and there are n elements in S,

$$P(S_i)=\frac{1}{n},\ i=1, 2, \ldots\ldots, n.$$

Let $\qquad A=\{S_1, S_2, \ldots\ldots, S_m\}.$

Then by definition,

$$P(A)=P(S_1)+P(S_2)+\ldots\ldots+P(S_m),$$

$$=\frac{1}{n}+\frac{1}{n}+\cdots\cdots+\frac{1}{n}, \text{ upto } m \text{ terms,}$$

$$=\frac{m}{n}.$$

Remark. $\quad P(A)=\dfrac{n(A)}{n(S)},$

where $n(A)$, and $n(S)$ denote the number of elements in A and S respectively.

Example 7. *What is the probability of throwing a number greater than 3, in a single throw of an unbiased die?*

Solution. The sample space of the experiment is

$$S=\{1, 2, 3, 4, 5, 6\}.$$

Let A be the event of throwing a number greater than three. Then

$$A=\{4, 5, 6\}.$$

Since the outcomes in S are equally likely,

$$P(A)=\frac{n(A)}{n(S)}=\frac{3}{6}=\frac{1}{2}.$$

Example 8. *A letter of the English alphabet is chosen at random. Calculate the probability that the letter so chosen*

(*i*) *is a vowel,*

(*ii*) *precedes m and is a vowel,*

(*iii*) *follows m and is a vowel.*

Solution. The sample space of the experiment is

$$S=\{a, b, c, d, \ldots\ldots, x, y, z\}, n(S)=26.$$

(*i*) Let V be the event that the letter chosen is a vowel.

Then $\quad V=\{a, e, i, o, u\}.$

Therefore,

$$P(V)=\frac{n(V)}{n(S)}=\frac{5}{26}.$$

(*ii*) Let A be the event that the letter precedes m and is a vowel ; then

$$A=\{a, e, i\}.$$

$$P(A)=\frac{n(A)}{n(S)}=\frac{3}{26}.$$

(*iii*) Let B be the event that the letter follows m and is a vowel ; then

$$B=\{o, u\}$$

$$P(B)=\frac{n(B)}{n(S)}=\frac{2}{26}=\frac{1}{13}.$$ **Ans.**

Example 9. *A red die and a green die (both) are thrown. What is the probability that the sum of the numbers shown exceeds 8 ?*

Solution. The sample space S of the experiment is given by $S=\{(m, n) : 1 \leqslant m,\ n \leqslant 6\}$. The event A is given by

$$A=\{(3, 6), (4, 5), (4, 6), (5, 4), (5, 5),$$
$$(5, 6), (6, 3), (6, 4), (6, 5), (6, 6)\}.$$

$$P(A)=\frac{n(A)}{n(S)}=\frac{10}{36}=\frac{5}{18}.$$

Example 10. *A bag contains 4 white, 5 blue and 6 red beads, similar except for colour. What is the probability that two beads drawn from the bag at random are both red ?*

Solution. There are 15 beads in all. Two beads may be drawn in ${}^{15}C_2=105$ ways. The sample space has, therefore, 105 elements. All these 105 outcomes are obviously equally likely. Of the 6 red beads, two may be drawn in ${}^{6}C_2=15$ ways. The event A, therefore, consists of 15 sample points. On the assumption of equal likelihood,

$$P(A)=\frac{15}{105}=\frac{1}{7}.$$

Example 11. *An event A in a sample space $S=\{S_1, S_2, \cdots\cdots S_n\}$ consists of k sample points. What is the probability of the opposite event A ?*

Since $\bar{A}=S \sim A$, therefore, A contains $n-k$ sample points.

Hence

$$P(\bar{A})=\frac{(n-k)}{n}=1-\frac{k}{n}=1-P(\bar{A}).$$

Remarks 1. Note that $P(A)+P(\bar{A})=1$.

2. The ratio $P(A) : P(\bar{A})$ is often talked of as the odds that the event A will occur. In example 5 above, the odds that A will occur are $k/n : (n-k)/n$ or $k : n-k$. This fact is also expressed by saying that the odds are k to $n-k$ in favour of A. Also, the odds that A will not occur are $(n-k)/n : k/n$ or $n-k : k$. This is expressed by saying that the odds are $n-k$ to k against A or that odds against A are $n-k : k$.

Example 12. *What are the odds in favour of drawing a king from a standard deck of 52 cards ?*

Solution. Let A be the event that a king is drawn. On the assumption that every card has the same likelihood of being drawn, the sample space contains 52 equally likely sample points. The event A contains 4 of these. Hence

$$P(A)=\frac{4}{52}=\frac{1}{13}.$$

$$P(\bar{A})=1-\frac{1}{13}=\frac{12}{13}.$$

The odds in favour of drawing a king are $\frac{1}{13}:\frac{12}{13}$ or 1 : 12.

EXERCISE 11 (*c*)

(Assume equally likely outcomes in all cases and do the first four exercises orally).

1. A single die is thrown. What is the probability of getting :
   (*a*) '5'. (*b*) '2' or '4'. (*c*) an odd number. (*d*) a multiple of three. (*A.I.S.S.C.E., 1986*)

2. Two coins are tossed. What is the probability of getting :
   (*a*) two heads, (*b*) at least one head, (*c*) no head.

3 The dial of a spinner for a game is divided into 12 equal sections as shown. Assuming the pointer does not stop on a dividing line, find the probability of spinning

   (*a*) a number greater than 2.

   (*b*) a number greater than or equal to 4.

   (*c*) the number 4.

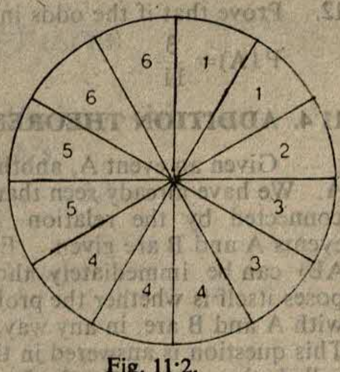


Fig. 11·2.

4. A card is drawn at random from a well-shuffled deck of 52 cards. Find the probability that it is neither an ace nor a king. (*D.B.S.S.C.E., 1989*)

5. In a single throw of two dice, what is the probability of obtaining a total of
   (*a*) 9 or 11. (*A.I.S.S.C.E., 1989*)
   (*b*) more than 10. (*D.B.S.S.C.E., 1988*)

6. A die is thrown twice. What is the probability of getting a sum greater than nine ?

7. A bag contains 4 white balls, 3 black balls and 5 red balls, similar except for colour. Three balls are drawn at random. What is the probability that :

(*a*) they are all white ? (*b*) they are all black ? (*c*) none of them is black ? (*d*) only one of them is white ? (*e*) at least one of them is white ? (*f*) all of them are of different colours ?

8. Two cards are drawn from a deck of 54 playing cards, two of them being jokers. What is the probability that :

(*a*) they are both spades ? (*b*) they are both kings ? (*c*) at least one of them is a joker ? (*d*) one of them is a queen and the other either a jack or a joker ?

9. What is the probability that a family with two children has one male and one female child, assuming the probability of each boy and girl to be $\frac{1}{2}$? [Think carefully ; the answer is not $\frac{1}{3}$.]

10. What are the odds against drawing a face card (king, queen, jack) from a standard deck of 52 cards ?

11. The probability of the happening of an event is p. What are the odds

(*a*) in its favour ? (*b*) against it ?

12. Prove that if the odds in favour of an event A are 3 : 8, then $P(A)=\frac{3}{11}$.

11·4. ADDITION THEOREM

Given an event A, another event that naturally arises is $\sim A$ or $\bar{A}$. We have already seen that the probabilities of these events are connected by the relation $P(A)+P(\bar{A})=1$. Suppose now two events A and B are given. Events $A\cup B$ (or $A+B$) and $A\cap B$ (or AB) can be immediately thought of. A natural question which poses itself is whether the probabilities of these new events connected with A and B are, in any way, connected with those of A and B. This question is answered in the affirmative in the following theorem called the **generalized theorem of total probability** or the **generalized theorem of addition of probabilities.**

Theorem. *For any event A, B subsets of a sample space S, $P(A\cup B)=P(A)+P(B)-P(A\cap B)$.*

Proof. If $A\cap B\neq\phi$, let $A\cap B=\{c_1, c_2, \ldots\ldots, c_s\}$. Then A and B may be written as

$$A=\{a_1, a_2, \ldots\ldots, a_q, c_1, c_2, \ldots\ldots, c_s\},$$
$$B=\{b_1, b_2, \ldots\ldots, b_r, c_1, c_2, \ldots\ldots, c_s\}.$$

all the a_i's being distinct from all the b_j's.

$$A \cup B = \{a_1, a_2, \cdots\cdots, a_q, b_1, b_2, \cdots\cdots, b_r, c_1, c_2, \cdots\cdots, c_s\}.$$

Now $\quad P(A) = \sum_{i=1}^{q} P(a_i) + \sum_{i=1}^{s} P(c_i).$

$$P(B) = \sum_{i=1}^{r} P(b_i) + \sum_{i=1}^{s} P(c_i),$$

$$P(A) + P(B) = \sum_{i=1}^{q} P(a_i) + \sum_{i=1}^{r} P(b_i) + \sum_{i=1}^{s} P(c_i) + \sum_{i=1}^{s} P(c_i),$$

$$= P(\{a_1, a_2, \cdots\cdots, a_q, b_1, b_2, \cdots\cdots, b_r, c_1, c_2, \cdots\cdots, c_s\}) + \sum_{i=1}^{n} P(c_i),$$

$$= P(A \cup B) + P(A \cap B),$$

$\therefore \; P(A \cup B) = P(A) + P(B) - P(A \cap B).$

If $A \cap B = \phi$, then let

$$A = \{a_1, a_2, \ldots\ldots, a_q\},\ B = \{b_1, b_2, \ldots\ldots, b_r\},$$

where the a_i's are distinct from the b_j's.

$\therefore \; A \cup B = \{a_1, a_2, \ldots\ldots, a_q, b_1, b_2, \cdots\cdots, b_r\}.$

and $P(A \cup B) = \sum_{i=1}^{q} P(a_i) + \sum_{i=1}^{s} P(b_i),$

$$= P(A) + P(B).$$

Since $A \cap B = \phi$, therefore, $P(A \cap B) = 0$.

Hence the above relation may again be written as

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

This proves the theorem for all events A and B.

Corollary 1. *If A and B are mutually exclusive events, then*

$$P(A \cup B) = P(A) + P(B).$$

Proof. Recall that A and B are said to be mutually exclusive if $A \cap B = \phi$.

Remark. The above corollary is known as the **theorem of total probability** or the **theorem of addition of probabilities.**

Corollary 2. *If A and B are mutually exclusive and exhaustive, then*

$$P(A)+P(B)=1.$$

Proof. $P(A)+P(B)=P(A\cup B)$, [$\because$ A and B are mutually exclusive.]

$=P(S)$, [$\because$ A and B are exhaustive.]

$=1.$

Remark. The above corollary is nothing but our relation

$$P(A)+P(\bar{A})=1$$

in a disguised form. Why ?

Corollary 3. *If $A_1, A_2, \ldots\ldots, A_n$ are events in a sample space any two of which are mutually exclusive, then*

$$P(A_1\cup A_2\cup\ldots\ldots\cup A_n)=P(A_1)+P(A_2)+\ldots\ldots+P(A_n).$$

Proof. Follows from the associativity of set union and principle of mathematical induction.

Aliter. Note that the sample points constituting $A_1,\ldots\ldots,A_n$ are all distinct. Then use the definition of the probability of an event.

Example 13. *A card is drawn at random from a well shuffled pack of 52 cards. What is the probability of getting a two of hearts or a two of diamonds ?*

Solution. Let 4 be the event of getting a two of hearts and B the event of getting a two of diamonds. Then clearly, $A\cap B=\phi$. The event "two of hearts or two of diamonds" is $A\cup B$.

Also, $P(A)=\dfrac{1}{52}$, $P(B)=\dfrac{1}{52}$.

$$\therefore\ P(A\cup B)=P(A)+P(B)=\frac{1}{52}+\frac{1}{52}=\frac{1}{26}.$$

Example 14. *A natural number is chosen at random from amongst the first 500. What is the probability that the number so chosen is divisible by 3 or 5.*

Solution. Let S be the sample space. Then

$S=\{1, 2, 3,\ldots\ldots,500\}$, $n(S)=500$.

Let A be the event that the number chosen is divisible by 3.

then $A=\{3, 6, 9,\ldots\ldots,498\}$, $n(A)=166$.

Let B be the event that the number chosen is divisible by 5.

Then $B=\{5, 10, 15,\ldots\ldots,500\}$, $n(B)=100$.

Now

$$P(A)=\frac{166}{500}=\cdot 332.$$

$$P(B)=\frac{100}{500}=\cdot 2.$$

Also, $A\cap B=\{15, 30, 45, \ldots\ldots, 495\}$, $n(A\cap B)=33$.

$$\therefore \quad P(A\cap B)=\frac{33}{500}=\cdot 066.$$

$$\begin{aligned}\therefore \quad P(A\cup B)&=P(A)+P(B)-P(A\cap B),\\ &=\cdot 332+\cdot 2-\cdot 066,\\ &=\cdot 532-\cdot 066,\\ &=\cdot 466.\end{aligned}$$

Example 15. *A batch of 15 radio tubes contains two defective tubes, which cannot be distinguished by looking merely. Two tubes are chosen at random from these 15. What is the probability that at least one of these is defective* ?

Solution. The sample space contains ${}^{15}C_2=105$ elements. The event A, *viz. at least one selected tube is defective* may be thought of as complementary of the event B, *viz. both tubes are acceptable*.

$$\begin{aligned}\therefore \quad P(A)&=P(\bar{B}),\\ &=1-P(B),\\ &=1-\frac{{}^{13}C_2}{105},\\ &=1-\frac{78}{105},\\ &=\frac{27}{105},\\ &=\frac{9}{35}.\end{aligned}$$

Example 16. *There are three events A, B, C, one and only one of which must accur. The odds are 8 to 3 against A and 5 : 2 against B. What are the odds against C* ?

Solution. Let S be the sample space. Then,

$$1=P(S)=P(A)+P(B)+P(C).$$

Now $\quad P(A)=\dfrac{3}{8+3}=\dfrac{3}{11},$

$$P(B)=\frac{2}{5+2}=\frac{2}{7}.$$

$\therefore \qquad P(C)=1-\frac{3}{11}-\frac{2}{7}=\frac{77-21-22}{77}=\frac{34}{77}.$

Odds against C are 43 : 34.

EXERCISE 11 (*d*)

1. Let A and B be events in a sample sapce S. Complete the following :

(*a*) If $P(A)=0{\cdot}3$, $P(B)=0{\cdot}4$, $P(A\cap B)=0{\cdot}1$, then $P(A\cup B)$ $=$.........

(*b*) If $P(A)=0{\cdot}4$, $P(B)=0{\cdot}5$, $P(A\cup B)=0{\cdot}7$, then $P(A\cap B)$ $=$.........

(*c*) If $P(A\cup B)=0{\cdot}8$, $P(A\cap B)=0{\cdot}2$, $P(A)=0{\cdot}5$, then $P(B)$ $=$.........

(*d*) If $P(A)=0{\cdot}3$, $P(B)=0{\cdot}2$, $P(A\cap B)=0{\cdot}1$, then $P(A'\cap B)$ $=$.........

[**Hint :** $B=(A\cup A')\cap B=(A\cap B)\cup(A'\cap B)$.]

2. Let A and B be mutually exclusive events in a sample space, such that $P(A)=0{\cdot}4$, $P(B)=0{\cdot}3$. Find the probability of
(*a*) $A\cup B$, (*b*) A', (*c*) B',
(*d*) $A'\cap B'$, (*e*) $A'\cap B$, (*f*) $A\cap B'$.

3. In a single throw of two dice, find the probability of not getting the same number on both the dice. (*A.I.S.S.C E. 1986*)

[**Hint :** How about using $P(A)=1-P(\bar{A})$?]

4. A card is drawn at random from a well-shuffled deck of 52 cards. Find the probability of its being a spade or a king. (*A.I.S.S.C.E., 1984*)

5. A husband and wife appear in an interview for two vacancies in the same post. The probability of husband's selection is 1/3 and that of the wife's selection is 1/5. Find the probability that only one of them will be selected. (*D.B.S.S.C E., 1984*)

6. Two events A and B have probabilites $0{\cdot}25$ and $0{\cdot}50$ respectively. The probability that both A and B occur simultaneously is $0{\cdot}14$. Find the probability that neither A nor B occurs. (*I.I.T.J.E.E., 1980*)

7. The probability that at least one of the events A and B occurs is $0{\cdot}6$. If A and B occur simultaneously with probability $0{\cdot}2$, then evaluate $P(A)+P(B)$. (*I.I.T.J.E.E., 1987*)

[Surprised at the answer? Remember $P(A)+P(B)$ is a sum of two real numbers and though each of them is less than one, the sum nevertheless may be greater than one. Do not confuse it with $P(A+B)$].

8. In a class of 200 students, 125 take mathematics, 112 take Physics and 47 take both mathematics and Physics. If a student is selected at random from this class, what is the probability that the student chosen takes mathematics or Physics ?

9. Six people seat themselves at random around a circular table. What is the probability that a given two of them are not together ?

10. A diminished deck of cards is formed from an ordinary deck of playing cards by removing all cards numbered less than nine (including aces). What is the probability that a card drawn at random from the diminished deck is either

 (*a*) a jack or a club,

 (*b*) either a face card or a spade.

11·5. CONDITIONAL PROBABILITY AND MULTIPLICATION THEOREM

Suppose an integer is chosen at random from amongst the first 1000 natural numbers. A sample space for this experiment is

$$S=\{1, 2, 3, \ldots\ldots, 1000\}\ ;\ n(S)=1000.$$

Let A and B respectively be the events that the selected integer is divisible by 5 and 2. Then

$$A=\{5, 10, 15, \ldots\ldots\ldots, 1000\}\ ;\ n(A)=200,$$

$$B=\{2, 4, 6, \ldots\ldots\ldots, 1000\}\ ;\quad n(B)=500.$$

How to find $A\cap B$? Isn't that easy ? We might start picking from S, the elements which belong to both of A and B. But we are sure that you are going to suggest an improvement. It is not really necessary to test *each* element of S. Since an element of S cannot belong to $A\cap B$ unless it belongs to both of A and B, we may start with the elements of A and examine which ones of these are in B too (or start looking up the elements of B and see which of them belong to A as well !). This reduces our labour.

In order to find $P(A\cap B)$, we have to explore the sample points which constitute $A\cap B$. All of these are included in A. We may forget about the points other than these. So we confine our attention to points of A. That is, we may regard A as our new sample space and see which of the sample points in A belong to B. $A\cap B\subset A$ and the *probability that a point of A chosen at random is also in* $B=n(A\cap B)/n(A)$. Be warned that this is not the same as $P(A\cap B)$—*the probability that a point of S selected at random is in A and on this assumption, we are calculating the probabitity of its being in B also.* In the latter case, no such assumption is made and we are simply calculating the probability that the element selected from S is in both A and B ; no doubt this latter probability is

$n(A\cap B)/n(S)$. The first probability is called the *conditional probability of event B on the assumption that A has occurred* and is denoted by $P(B \mid A)$. It is related with $P(A\cap B)$ in the following manner :

$$P(B \mid A)=\frac{n(A\cap B)}{n(A)},$$

$$=\frac{n(A\cap B)}{n(S)}\Big/\frac{n(A)}{n(S)}, \text{ assuming S to be finite,}$$

$$=\frac{P(A\cap B)}{P(A)}$$

We formalize the above idea in the following definition :

Definition. *Let A and B be two events of a sample space S, with probability function P. Then the conditional probability of B, given that A has occurred, is denoted by $P(B \mid A)$, and is given by*

$$P(B \mid A)=\frac{P(B\cap A)}{P(A)}, \qquad P(A)\neq 0.$$

If $P(A)=0$, then $P(B \mid A)$ is not defined.

Conditional probabilities play another role also. It is a matter of common experience that if two events A, B say, can occur as a result of a random experiment, then the probability of happening of one of them say B, in some cases, is influenced by the occurrence or non-occurrence of A. The extreme cases are (*i*) B always occurs if A does, and (*ii*) B never occurs if A does. To measure the extent to which the occurrence of one is influenced (or *not* influenced at all !) by the occurrence of the other, the concept of conditional probability of one on the assumption of the other having happened is found to be useful.

Example 17. *Six cards—ace, king and queen of hearts ; ace and king of diamonds ; and king of clubs are placed face down on the table. A card is drawn at random. Given that a heart has been selected, what is the probability that it is a king ?*

Solution. The sample space of the experiment is

$$S=\{H_a,H_k,H_q,D_a,D_k,C_k\},$$

H_a denoting the ace of hearts and so on.

Let H be the event that a 'heart' is drawn and K the event that a 'king' is drawn. Then

$$P(H)=\frac{3}{6}=\frac{1}{2},$$

$$P(K)=\frac{3}{6}=\frac{1}{2}$$

$$P(H\cap K)=\frac{1}{6}$$

$$P(K \mid H)=\frac{P(H\cap K)}{P(H)}=\frac{1/6}{1/2}=\frac{1}{3}.$$

As stated earlier, our interest in conditional probabilities lies because they help us in calculating the probabilities of joint events $A\cap B$. The following theorem known as the *theorem of compound probability* or the multiplication theorem of probobilities justifies this claim.

Theorem. (*Multiplication theorem of Probabilities*) *Let A and B be any events of a sample space S, with probability function P such that $P(A)\neq 0$ and $P(B)\neq 0$. Then*

$$P(A\cap B)=P(B)P(A \mid B)=P(A)P(B \mid A).$$

Proof. Since $P(B)\neq 0$, we have

$$P(A \mid B)=\frac{P(A\cap B)}{P(B)}$$

$$\therefore \quad P(A\cap B)=P(B)P(A \mid B).$$

Similarly,

$$P(A\cap B)=P(A)P(B \mid A).$$

Remark. In most cases, it is easy to calculate $P(B \mid A)$ directly. In such cases, the above theorem gives a formula for calculating $P(A\cap B)$.

Example 18. *A card is drawn from a well shuffled pack of 52 cards and the outcome noted. Another card is now drawn from the pack, without replacing the first. What is the probability that the first card is a king and the second a queen of a different suit ?*

Solution. Let the events A and B be :

A=first card is a king,

B=the second card is a queen of a different suit.

Then $A\cap B$=the first card is a king and the second a queen of a different suit.

$$P(A)=\frac{4}{52}=\frac{1}{13}.$$

If we assume that A has occurred, then 51 cards are left which contain all the 4 queens and therefore three queens of suits different from that of the king drawn.

$$\therefore \quad P(B \mid A)=\frac{3}{51}=\frac{1}{17}$$

$$P(A\cap B)=P(A)P(B \mid A).$$

$$=\frac{1}{13}\cdot\frac{1}{17}.$$

$$=\frac{1}{221}.$$

Example 19. *An urn contains 7 red and 4 blue balls. Two balls are drawn at random, without replacement. What is the probability that both the balls are red?*

Solution. Consider the events

A = the first ball is red, and

B = the second ball is red.

Then $A \cap B$ = both balls drawn are red.

Then we know

$$P(A)=\frac{7}{11}.$$

Also, $P(B \mid A)=\frac{6}{10}=\frac{3}{5}$ for if A occurs, then we are left with 10 balls, 6 of which are red.

$$\therefore \quad P(A\cap B)=P(A)P(B \mid A),$$

$$=\frac{7}{11}\cdot\frac{3}{5},$$

$$=\frac{21}{55}.$$

EXERCISE 11 (*e*)

1. In a group of 100 persons, 30 are lawyers, 70 are liars and 25 are both. One out of these 100 persons is selected at random. The events A and B are respectively defined as

   A = selected person is a lawyer,

   B = selected person is a liar.

   Calculate the following probabilities and mention which of them, if at all, are $P(A \mid B)$, $P(B \mid A)$?

   (*a*) Probability that the selected person is a lawyer if he is known to be a liar.

   (*b*) Probability that the selected person is a liar if he is known to be a lawyer.

   (*c*) Probability that the selected person is both a lawyer and a lair.

2. In a random experiment of selecting a number from amongst the first 10,000 natural numbers, calculate the probability that the selected number is

   (*a*) divisible by 5.

   (*b*) divisible by 5 if it is known to be an even number. (Is the result surprising?)

   (*c*) divisible by 10.

(*d*) divisible by 10 if it is known to be divisible by 4. (Do not jump to conclusions ! The calculation might fail your intuition !).

3. A bag contains 8 green and 10 white balls. Two balls are drawn. What is the probability that one is green and the other is white ? (*D.B.S.S.C.E., 1988*)

4. A bag contains 5 green and 7 red balls. Two balls are drawn. What is the probability that one is green and the other is red ? (*A.I.S.S.C.E 1984*)

5. Two cards are drawn from a well-shuffled pack of 52 cards, one after another without replacement. Find the probability that one of these is an ace and the other is a queen of the opposite shade. (*D.B.S.S.C.E, 1985*)

6. Two cards are drawn one after another from a pack of 52 ordinary cards. Find the probability that the first card drawn is an ace and the second is an honour (Ace or King or Queen or Jack) card. The first card is not replaced while drawing the second. (*Roorkee, 1983*)

7. A bag contains 19 tickets, numbered from 1 to 19. A ticket is drawn and then another ticket is drawn without replacement. Find the probability that both tickets will show even numbers. (*A.I.S.S.C.E., 1986*)

8. A vendor mixes 50 balls of second quality with 100 similar-looking balls of first quality. A child picks up two balls at random. What is the probability that

   (*a*) both the balls are of first quality ?

   (*b*) exactly one ball is of first quality ?

9. A bag contains 5 white, 7 red and 8 black balls. If four balls are drawn one by one without replacement, what is the probability that all are white ? (*A.I.S.S.C.E, 1987*)

10. A bag contains 5 white and 3 black balls. 4 balls are drawn one at a time without replacement. Find the probability that the balls are alternately of different colours. (*D.B.S.S.C.E. 1989, A.I.S.S.C.E. 1988*)

11. Prove that :

    (*a*) $0 \leqslant P(A \mid B) \leqslant 1$.

    (*b*) If A and B are mutually exclusive, then $P(A \mid B) = 0$.

    (*c*) If $B \subset A$, then $P(A \mid B) = 1$.

11.6. INDEPENDENT EVENTS

Consider a random experiment of throwing a red and a green die. It is obvious that the occurrence of a certain number of dots on the red die has nothing to do with a similar event for the green die. The two are quite independent of each other, so to say. But

suppose, the two dice were tied to the two ends of a piece of thread before being thrown. The situation changes. This time the two events are not independent in as much as the uppermost face of one die will have something to do in causing a particular face of the other die to be uppermost ; and the shorter the thread, the more is this influence or dependence. There are ways in which the degree of this dependence can be measured ; but at the moment we are interested in a very special case of dependence and that is the case of *no dependence* or *independence.* Suppose that the happening or otherwise of an event A has no influence on the probability of the occurrence of another event B in the same sample space S. Then probability of B remains the same whether or not A occurs. This means that probability of B is the same as the probability of B on the assumption that A has occurred or is sure to occur. This amounts to saying that

$$P(B)=P(B \mid A).$$

This suggests the following definition :

Definition. Let A and B be two events of a sample space S. The event B is said to be **independent of** A if

$$P(A)\neq 0 \text{ and } P(B)=P(B \mid A).$$

Theorem. *If A is independent of B, and $P(A)\neq 0$, then B is also independent of A.*

Proof. $\because$ A is independent of B,

$$P(B)\neq 0 \text{ and } P(A)=P(A \mid B),$$

i.e., $$P(A)=\frac{P(A\cap B)}{P(B)},$$

i.e., $$P(B)=\frac{P(A\cap B)}{P(A)}, \quad (\text{given } P(A)\neq 0)$$

i.e., $$P(B)=P(B \mid A),$$

i.e., B is independent of A.

Corollary. *In case both P(A) and P(B) are positive, an event A is independent of another event B on the same sample space if and only if B is independent of A.*

This symmetry enables us to talk about the two events being independent rather than one being independent of the other. The multiplication theorem of probabilities assumes a very simple and pleasing form in the case of independent events, as the reader must already have realized. We state it for the sake of completeness.

Theorem. *If A and B are independent events defined on the same sample space, then*

$$P(A\cap B)=P(A)P(B).$$

Proof. $$P(A\cap B)=P(A)\,P(B \mid A)$$
$$=P(A)\,P(B).$$

Remarks 1. Independent events have positive probability.

2. Any of the following three may be taken as a criterion (criteria is the plural please !) of A and B being independent :

(*i*) $P(A) = P(A \mid B)$.

(*ii*) $P(B) = P(B \mid A)$.

(*iii*) $P(A \cap B) = P(A)\, P(B)$.

3. We may talk of more than two events being independent, as follows :

Definition. *The events $A_1, A_2, \ldots\ldots, A_n$ defined on the same sample space with probability function P, are said to be independent if*

$$P(A_i \cap A_j) = P(A_i)\, P(A_j),$$
$$P(A_i \cap A_j \cap A_k) = P(A_i)\, P(A_j)\, P(A_k),$$
$$\vdots \qquad\qquad \vdots$$
$$P(A_1 \cap A_2 \cap \ldots\ldots \cap A_n) = P(A_1)\, P(A_2) \ldots\ldots P(A_n),$$

for all combinations of indices such that $1 \leqslant i < j < k \leqslant n$.

Thus if a set of events is independent, then any two of them are independent but the converse is not true as shown in Example 25 further on.

Example 20. *A box contains five white balls and seven black balls. Two balls are drawn in succession. What is the probability that both are white if the first ball is replaced before the next one is drawn ?*

Solution. Let A be the event that a white ball is drawn on the first draw. Then

$$P(A) = \frac{5}{12}.$$

Let B be the event that a white ball is drawn on the second draw. We then wish to find the probability of $A \cap B$.

Also, since the ball drawn has been replaced before the second is drawn, the occurrence of B is not affected by the colour of the first ball drawn, *i.e.*, whether or not the first ball is white,

$P(B) = \frac{5}{12}$, and A, B are independent.

$$P(A \cap B) = P(A)\, P(B) = \frac{5}{12} \cdot \frac{5}{12} = \frac{25}{144}.$$

Hence the probability that the two balls drawn are both white

$$= \frac{25}{144}. \qquad \textbf{Ans.}$$

Example 21. *Find the probability of throwing at least one 1 in two throws with one die.*

Solution. Let A be the event of throwing a 1 on 1st throw and B the event of throwing a 1 on the 2nd throw. Then

$$P(A)=\frac{1}{6},\ P(B)=\frac{1}{6}.$$

Clearly, A and B are independent of each other. Hence A′ and B′ are also independent of each other.

Now, if C be the event of throwing at least one 1 in two throws, then

$$C=A\cup B.$$
$$C'=A'\cap B'.$$

Since $$P(A')=1-\frac{1}{6}=\frac{5}{6},\ P(B')=\frac{5}{6},$$

$$\therefore\quad P(A'\cap B')=P(A')\cdot P(B')=\frac{25}{36},\ \text{or}\ P(C')=\frac{25}{36}.$$

Hence $$P(C)=1-P(C')=1-\frac{25}{36}=\frac{11}{36}.$$

We now consider some historical examples of the same nature as Example 2 above, but solve them by a different technique.

Example 22. (*Problem of repeated birthdays*) *Assuming that each person present in a room can have as his birthday any of the 365 days of the year and that all the days of the year are equally likely to be his birthday, what is the probability that at least two of the n persons present in the room will have a common birthday?*

Solution. We observe first of all that the events A=*at least* two persons have a common birthday, and B=*no* two persons have a common birthday, are complementary of each other so that

$$P(A)+P(B)=1.\qquad [B=\overline{A}]$$
$$\therefore\qquad P(A)=1-P(B).$$

Let us name the persons in the room as $P_1, P_2, \ldots\ldots,$ and P_n, and let the events A_i be defined as follows:

$A_i=P_i$ has a birthday different from $P_1, P_2, \ldots\ldots\ldots, P_{i-1}$, $2 \leqslant j \leqslant n$.

Then $$B=A_2\cap A_3\cap\ldots\ldots\cap A_n.$$

The events $A_2, A_3, \ldots\ldots, A_n$ are independent because of the hypotheses and hence

$$P(B)=P(A_2)\ldots\ldots P(A_n).$$

Now P_1 has a certain birthday. In order that P_2 may have a birthday different from that of P_1, his birthday should fall on one of the remaining 364 days of the year and the probability of this is

$$\frac{364}{365}=1-\frac{1}{365}.$$

Thus $$P(A_2)=1-\frac{1}{365}.$$

Similarly $P(A_3)$, the probability of P_3 having a birthday different from that of P_1 and P_2 both, is $\frac{363}{365}=1-\frac{2}{365}$ and so on so forth, till we get $P(A_n)=1-\frac{n-1}{365}$.

Hence, $$P(B)=P(A_2)\cap\cdots\cdots\cap P(A_n),=\left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\cdots\cdots\left(1-\frac{n-1}{365}\right)$$

$$\therefore \quad P(A)=1-P(B),$$
$$=1-\left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\cdots\cdots\cdots\left(1-\frac{n-1}{365}\right).$$

Try to guess how many persons must be present in a room so that the probability of at least two persons having a common birthday may exceed 0·5. Now look at the following table giving n (the number of persons in the room) and P (A) (the probability of at least two of these n persons having a common birthday).

n	P (A)
4	0 016
8	0 074
12	0·167
16	0·284
20	0 411
21	0·444
22	0·476
23	0·507
24	0·538
30	0 706
60	0·994

How far was your guess from the correct answer 23 ? The following is another example where instuition fails us (or at least it failed the famous diarist Samuel Pepys !)

Example 23. *In the year 1693, Pepys expressed his wonderment in a letter to Sir Isaac Newton on the following oddity :*

A man asserts that he will throw at least one six by throwing 6 dice once. Another man asserts that he will throw at least two sixes by throwing 12 dice once. Yet another man asserts that he will throw at least three sixes by throwing 18 dice once.

It is surprising that the probabilities of the fulfilment of their promises are not the same.

Let us calculate first the probability of the first man's assertion being proved true.

The events R—"at least one six" and Q—"no six", are complementary. The probability of a particular die not showing a six is $\frac{5}{6}$ no matter what the other dice show. Thus the events of the dice not showing a 6 are independent. The event *no die shows a six* is the intersection of the six events that the 6 individual dice do not show a six. Hence

$$P(Q)=\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}$$

$$P(R)=1-\left(\frac{5}{6}\right)^6=\frac{31031}{46656}=0{\cdot}665.$$

Now let us consider the probability of the assertion of the second man being true. The probability of at least two sixes in 1 throw of 12 dice may be found by considering the following events :

A=at least 2 sixes

B=exactly 1 six

C=no six.

A, B and C are pairwise mutually exclusive (*i.e.*, any two of them are mutually exclusive) and exhaustive.

$$P(A)+P(B)+P(C)=1.$$

or $$P(A)=1-P(B)-P(C).$$

Now $P(C)=\left(\frac{5}{6}\right)^{12}$, using arguments similar to those used in the previous case.

For the event B, exactly one die is to show a six and the others are to show a non-six.

This one die may be any one of the twelve dice. If we denote by B_j the event in which the jth dice shows a six and the others do not, $j=1, 2, \ldots\ldots, 12$, then

$$B=\cup\,\{B_j : j=1, 2, \ldots\ldots, 12\} \qquad \ldots(1)$$

Now if the ith die shows a six and exactly one die is to show a six, then for $j\neq i$, the jth die cannot show a six. Therefore, B_i and B_j are mutually exclusive if $i\neq j$. Thus in view of (1),

$$P(B)=\sum_{j=1}^{12} P(B_j). \qquad \ldots(2)$$

Let us calculate $P(B_j)$ for any $j=1, 2, \ldots\ldots, 12$. Let C_i be the event that the ith die D_i shows a six, $i=1, 2, \ldots\ldots, 12$. Then $\overline{C}_i$ is the event that D_i shows one of 1, 2, 3, 4 and 5. Hence $P(C_i)=\frac{1}{6}$ and $P(\overline{C}_i)=\frac{5}{6}$. The event B_j can be written as

$$B_j=\overline{C_1}\ \overline{C_2}\cdots\cdots\overline{C}_j-1\ \overline{C}_j\ .\ \overline{C}_j+.\ \cdots\cdots\overline{C}_{12}$$

The events of various dice showing a particular digit are independent.

Hence

$$P(B_j)=P(C_1)\,P(C_2)\ldots\ldots P(C_{j-1})\,P(C_j)\,P(C_{j+1})\cdots\cdots P(C_{12})$$

$$=\frac{5}{6}\cdot\frac{5}{6}\cdots\cdots\cdots\frac{5}{6}\cdot\frac{1}{6}\cdot\frac{5}{6}\cdots\cdots\cdots\frac{5}{6}=\frac{1}{6}\left(\frac{5}{6}\right)^{11}.$$

The equation (2) above now gives

$$P(B)=\frac{1}{6}\cdot\left(\frac{5}{6}\right)^{11}+\frac{1}{6}\cdot\left(\frac{5}{6}\right)^{11}+\ldots\ldots 12 \text{ times},$$

$$=12\,.\,\frac{1}{6}\cdot\left(\frac{5}{6}\right)^{11}.$$

$$P(A)=1-12\,.\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{11}-\left(\frac{5}{6}\right)^{12},$$

$$=\frac{1346704211}{2176782336},$$

$$=0{\cdot}619<0{\cdot}665.$$

Hence the probability of the assertion of the second man being true is lesser than that of the first. It can be shown similarly that the third man has still lesser probability, namely 0·597, of fulfilling his promise.

Remark. In example 23 above, we considered compound events which were intersection or union of 6 or 12 other events. In such complicated situations, probability trees are very useful. We shall illustrate the idea of a probability tree by means of an example.

Example 24. *A box contains 3 red and 2 green balls. Two balls in succession are drawn without replacement. What is the probability that (i) both balls are red? (ii) first ball is green and the second red?*

Solution. At the first draw, either a red (R) or a green (G) ball may be drawn and the probabilities of these two events are respectively 3/5 and 2/5. We graph this situation in the figure overleaf by the first two branches of the tree starting at 0. The second ball drawn may again be red or green but the probability of its being red or green depends upon the colour of the first ball drawn. Suppose the first ball drawn was red, which is depicted by the

upper branch starting at 0. The conditional probabilities of a red and a green ball on the assumption of first ball being red are 2/4 and 2/4 because we are left with 4 balls 2 of which are red and 2 of which are green. These events and the corresponding conditional probabilities are depicted by the two sub-branches starting at R in the figure overleaf. The top-most branch ORR of the figure represents the event that both the balls drawn are red. The probability of this event is obtained by multiplying the probabilities listed on this branch (as well you know by now, why !) and is thus 3/10. The other part of the figure may now be interpreted in a similar light and the other required probability given by the branch OGR is also 3/10.

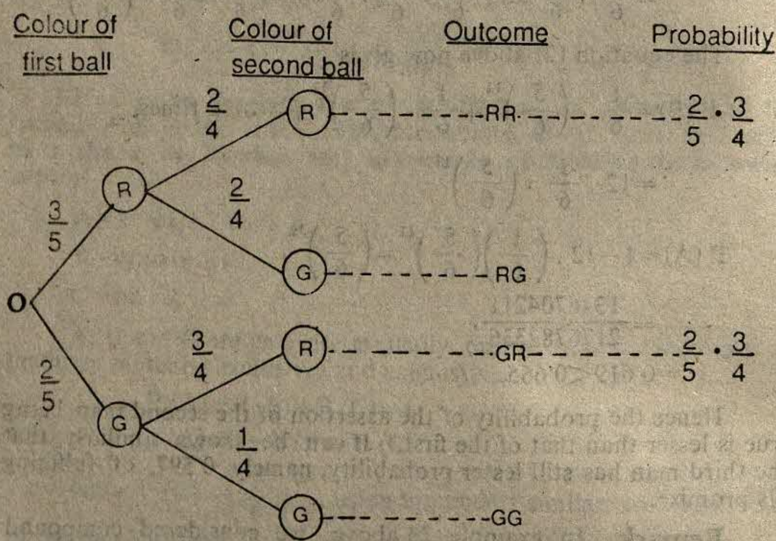


Fig. 11·3.

Remarks 1. The terminal points (the points from which no branch is going out) of the tree are really the sample points. But be careful. They are not equally likely. Had we numbered the balls R_1, R_2, R_3, G_1, G_2 and started with a five-branch tree instead of two etc. the final sample points would have been equally likely and at every stage each branch will have the same probability listed and we would have to count the number of sample points in the required event etc. and it would be more cumbersome than the algebraic proof !

2. Instead of drawing your tree from left to right, you may draw your tree top to bottom if you find that more convenient.

Example 25. *Consider the random experiment of throwing an unbiased tetrahedral die (a four-faced die like a triangular pyramid*

on a triangular base. A sample space for this experiment is {1, 2, 3, 4}. Let the events A, B and C be defined as :

$$A=\{1, 2\},\ B=\{1, 3\},\ C=\{1, 4\}.$$

Then $A \cap B=A \cap C=B \cap C=\{1\}=A \cap B \cap C$.

Also, $P(A)=P(B)=P(C)=\frac{1}{2}$

$$P(A \cap B)=P(A \cap C)=P(B \cap C)=\tfrac{1}{4}=P(A)P(B)$$
$$=P(A)P(C)=P(B)P(C).$$

Thus, $P(A B)=P(A)P(B)$, etc.

This shows that A and B are independent ; A and C are independent and that B and C are independent.

But since $P(A \cap B \cap C)=P(1)=\frac{1}{4}\neq P(A)P(B)P(C)$, therefore A, B and C are not independent.

11·7. INDEPENDENT EXPERIMENTS

Sometimes we may be interested in events arising out of performing several experiments in some context. For some reason we may also be interested in repeating the same experiment over and over again. For example, you may want to keep on rolling a dice over and again till you get a six in order to start the games like Ludo or Snakes and Ladders. Another example is that of firing a number of shots at a target.

In real life situations, we may be interested in gaining some insight into the behaviour or characteristics of a large group of persons or objects. For this, studying the whole group may be impossible, not feasible, or very costly. We then settle for selecting a small sample and study it. From our findings about the sample, we draw conclusions about the whole group and also assign numerical measure to our assertions being true. This, by the way, is all that Statistics is about, and this process cannot be carried out without an understanding of the principles of probability. Selecting a sample amounts to picking up items from the group one by one. Choosing an item without any bias may be considered as a random experiment and we may have to repeat it.

In a series of experiments, let 0 be any outcome of any experiment. Now two types of situations arise. Firstly, it is possible that the probability of this outcome is affected by the outcomes of the experiments carried out earlier ; secondly, it may not be affected by the outcomes of the earlier experiments. For example, if our experiment is *tossing a fair coin*, then the probability of a head in a particular toss remains $\frac{1}{2}$, no matter what the outcomes of the earlier tosses. On the other hand, if our experiment is *drawing a card* from an ordinary well-shuffled pack without replacement, then the probability of a diamond in a particular draw is affected by the results of the earlier draws. Experiments of the first type are known as *independent experiments.*

Definition. *Experiments in a series of experiments are said to be independent if the probability of any outcome of any one of these experiments remains unaffected by the outcomes of the earlier experiments in the series.*

Remark. Independent experiments need not be confused with independent events. When we were taking about events, we had only one sample space and the events were subsets of this one sample space. In a series of experiments, each experiment has its own sample space and the various sample spaces need not be the same. However, our interest lies in series of experiments which are merely repetitions of one particular experiment, and have identical sample spaces. This would be done in the next chapter. There we shall see how to calculate the probabilities of events associated with a series of independent experiments. This is a very fundamental and fruitful activity and depends solely on two principles learnt in this chapter—the theorems of compound and total probability.

EXERCISE 11 (*f*)

1. Given two independent events A and B, such that $P(A)=0{\cdot}4$, $P(B)=0{\cdot}3$, determine :

   (*i*) $P(A \cap B)$ (*ii*) $P(A' \cap B)$

   (*iii*) $P(A \cup B)$ (*iv*) $P(A' \cap B')$.

2. A professional magician claims to be a *mind-reader.* To test his claim, numbers 1 to 10 are written on cards. A spectator selects one card and the magician *reads his mind* by telling the number. Another card is selected and the *mind read* again. Assuming that the magician is merely guessing and pronouncing a number at random, what is the probability that he would pronounce the correct number both the times if the first card is

   (*a*) replaced before the second is drawn.

   (*b*) not replaced before the second is drawn.

3. A school enters 3 students in various sport competitions, the probability of each winning a prize (there are to be 3 in all) being $0{\cdot}1$. Assuming the events of their winning a prize to be independent, find the probability that

   (*a*) at least 1 gets a prize

   (*b*) at least 2 get a prize

   (*c*) all three of them get a prize

   (*d*) none of them gets a prize.

4. Let A and B be two independent events such that the probability of the occurrence of both of them simultaneously is $\frac{1}{6}$ and the probability that neither of them will occur is $\frac{1}{3}$. Find the probabilities of the occurrence of A and B. Are they unique ?

[**Hint** : What is the event opposite to *neither of A and B will happen* ? Is it $A \cup B$? Now make use of addition (generalized) and multiplication theorems.]

5. Three persons enter a railway compartment in which only two window seats are vacant. They decide to occupy these seats by tossing a fair coin in the alphabetical order of their names. The first two to throw a head are to occupy the window seats. More than one round of tosses are agreed upon if necessary. Their names turn out to be Suresh, Naresh and Dinesh. What are the probabilities of their occupying a window seat on the basis of first round of tosses ?

6. An article manufactured by a company consists of two parts X and Y. In the process of manufacture of the part X, 9 out of 100 parts may be defective. Similarly, 5 out of 100 are likely to be defective in the manufacture of the part Y. Calculate the probability that the assembled product will not be defective. (*Roorkee Entrance, 1989*)

7. If four numbers taken at random are multiplied together, find the probability that the last digit in the product is 1, 3, 7 or 9.

   [**Hint** : It is sufficient to consider the last digit, *i.e.* the unit's digit. By *numbers* you must understand *integers* here, for otherwise the problem makes no sense. Can any of the numbers in the product be a multiple of 2 or 5 ?]

8. The probabilities of the occurrence of n independent events are $p_1, p_2, \cdots, p_n$. Show that the probability that *at least one* of them would occur is

$$1-(1-p_1)(1-p_2)\cdots\cdots(1-p_n).$$

9. A problem in mathematics is given to three students whose chances of solving it are 1/2, 1/3, 1/4. What is the probability that the problem will be solved ? (*A.I.S.S.C.E., 1985*)

10. Given the probability that A can solve a problem is 2/3 and the probability that B can solve the same problem is 3/5, find the probability that (*i*) at least one of A and B will be able to solve the problem, (*ii*) none of the two will be able to solve the problem. (*Roorkee Entrance, 1980*)

11. The probabilities of A, B, C solving a problem are 1/3, 2/7 and 3/8 respectively. If all the three try to solve the problem simultaneously, find the probability that exactly one of them will solve it. (*D.B.S.S.C.E., 1987*)

12. A bag contains 3 red and 5 black balls, and a second bag contains 6 red and 4 black balls. A ball is drawn from each bag. Find the probability that one is red and the other is black. (*A.I.S.S.C.E., 1986*)

13. A purse contains 2 silver and 3 copper coins. A second purse contains 4 silver and 3 copper coins. If a coin is pulled out at random from one of the two purses, what is the probability that it is a silver coin ? *(D B.S.S.C.E., 1986)*

14. Three groups of children contain respectively three girls and one boy, two girls and two boys, one girl and three boys. One child is selected at random from each group. What is the chance that the three selected consist of one girl and two boys ? *(Roorkee Entrance, 1985)*

15. A one-man jury has probability p of making a correct decision. Another jury consists of three members, two of whom make a correct decision independently with a probability p each. The third member flips a coin to make decision. Majority decision is declared. Show that each jury is as well as the other.

16. Eight coins are thrown simultaneously. Find the probability of getting at least 6 heads.

(A.I S S.C.E. 1985, D.B S.S.C.E. 1987)

[**Hint.** Use the technique of Example 23.]

17. Ten coins are tossed simultaneously. Find the probability of getting at least seven heads. *(D.B.S.S.C.E. 1985)*

TEST YOUR UNDERSTANDING XI

In each of the following problems, tick the correct alternative.

1. This test contains 12 multiple choice items, each having four responses only one of which is the correct one. If you start making the responses at random, the probability of your getting all correct responses is

(a) $\frac{1}{4}$ (b) $\frac{1}{12}$

(c) $\frac{1}{48}$ (d) $\frac{1}{4^{12}}$.

2. In problem 1, the probability of your getting all responses wrong is

(a) $\frac{3}{4}$ (b) $\frac{3}{12}$

(c) $12.\left(\frac{3}{4}\right)^{12}$ (d) $\left(\frac{3}{4}\right)^{12}$.

[Observe that the probability of all responses right is less than 0·0000001 and that of 6 responses wrong is less than ·05. What lesson does it teach you ? *Do not guess* ; *think hard rather*.]

3. On a sample space, consisting of 4 elements, the maximum number of distinct events that can be defined, is :

(*a*) 4 (*b*) 8
(*c*) 16 (*d*) 32.

4. A bag contains 2 red and 1 black ball. Two balls are taken out at random. A sample space for this random experiment

(*a*) may contain 3 but not 6 elements
(*b*) may contain 6 but not 3 elements
(*c*) can contain neither 3 nor 6 elements
(*d*) may contain 3 as also 6 elements.

5. An NCC cadet fires at a target like the one shown in the adjoining figure. Which is a possible set of probabilities of hitting the bull's eye (inner-most disk) and the two circular rings ?

(*a*) 0·21, 0·35, 0·45,
(*b*) 0·15, 0·40, 0·55.
(*c*) 0·15, 0·45, 0·41,
(*d*) 0·25, 0·35, 0·40.

[**Hint.** What two properties must every probability function satisfy ?]

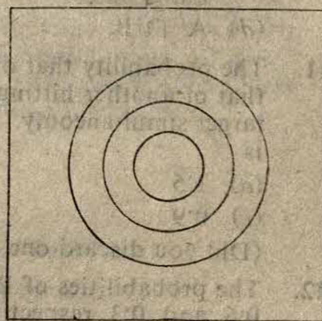


Fig. 11·4.

6. The probability, that a randomly selected natural number will have a 1 in the unit's place when squared, is equal to

(*a*) 0·2 (*b*) 0·4
(*c*) 0·04 (*d*) none of these.

[**Hint** : It is enough to consider one digit numbers.]

7. The letters of the word *TEACH* are written on cards, the cards are mixed and then rearranged. The probability that the letters now spell *CHEAT* is

(*a*) $\frac{1}{240}$ (*b*) $\frac{1}{120}$
(*c*) $\frac{1}{60}$ (*d*) $\frac{1}{3}$.

8. The letters of the word *STATES* are scrambled and rearranged. The probability of the rearranged word being *TASTES*, is

(*a*) $\frac{1}{720}$ (*b*) $\frac{1}{360}$
(*c*) $\frac{1}{180}$ (*d*) none of these.

[Careful ! There are 2 S's and 2T's. You had better use the multiplication theorem even if you used combinatorial methods in the previous problem.]

9. If $P(A \mid B) > P(A)$, then

(*a*) $P(B \mid A) < P(B)$ (*b*) $P(B \mid A) = P(B)$

(*c*) $P(B \mid A) < P(B)$ (*d*) none of (*a*), (*b*) and (*c*).

10. A positive integer is selected at random. Let A be the event that it is divisible by 5 and let B be the event that it has a zero at the unit's place. The event $A \cup B'$ is

(*a*) the impossible event (*b*) the sure event

(*c*) the event that the number has a non-zero digit at the unit's place

(*d*) $A' \cap B$.

11. The probability that one NCC cadet hits a target is 0·7 and that of another hitting the target is 0·8. They both fire at the target simultaneouly. The probability of the target being hit is

(*a*) 1·5 (*b*) 0·15

(*c*) 0·9 (*d*) none of these

(Did you discard one answer outright ?)

12. The probabilities of Ramesh and Shyam hitting a target are 0·6 and 0·3 respectively. The reward for hitting the target is Rs. 90. Both of them shoot at the target but only one of them hits the target. It is unknown as to who hit the target. In a fair distribution of reward, the share of Ramesh would be

(*a*) Rs. 45 (*b*) Rs. 60

(*c*) Rs. 70 (*d*) Rs. 90.

13. If the events A and B be mutually exclusive, then $P(A+B)$ will be equal to

(*i*) $P(A)+P(B)$ (*ii*) $P(A)-P(B)$

(*iii*) $P(A) \cdot P(B)$ (*iv*) $P(A)/P(B)$ (*MNR, 1978*)

14. A determinant is chosen at random from the set of all determinants of order 2 with elements 0 or 1 only. The probability that the value of the determinant chosen is positive is

(*a*) $\frac{1}{6}$ (*b*) $\frac{1}{8}$

(*c*) $\frac{3}{16}$ (*d*) 1.

15. A box contains 100 tickets numbered 1, 2,...,100. Two tickets are chosen at random. It is given that the maximum number on the two chosen tickets is not more than 10. The minimum number on them is 5 with probability

(a) ${}^{6}C_2 \,/\, {}^{100}C_2$ (b) ${}^{5}C_2 \,/\, {}^{100}C_2$

(c) ${}^{6}C_2 \,/\, {}^{10}C_2$ (d) ${}^{5}C_2 \,/\, {}^{10}C_2$.

16. If $\frac{1+3p}{3}$, $\frac{1-p}{4}$ and $\frac{1-2p}{2}$ are the probabilities of three mutually exclusive events, then the set of all values of p is

(a) $\{\frac{1}{3}\}$ (b) $[\frac{1}{3}, \frac{1}{2}]$

(c) $]\frac{1}{3}, \frac{1}{2}[$ (d) $[\frac{1}{3}, \frac{1}{2}[$

[**Hint.** Events need not be exhaustive. Use $0 \leqslant$ probability $\leqslant 1$.]

17. An urn contains a few white and a few black balls. The probability of the first drawn ball from the urn being white is p. A ball is drawn from the urn and without being observed for colour is put into a bag. Another ball is now drawn from the urn. The probability of this second ball being white

(a) is p

(b) is greater than p

(c) is less than p

(d) cannot be determined from the given data.

[**Hint.** How does it matter where the balls are !]

18. The term *reliability* is used for the *probability that a device does not fail.* A mechanical device D has two components which work independently. The reliabilities of these components are 0·92 and 0·95 respectively.

(i) If D can work only when both its components work, for example a two-battery torch, the reliability of

D is

(a) $\cdot 92 + 95$, (b) $\cdot 92 \times \cdot 95$,

(c) $\cdot 08 \times \cdot 05$, (d) $\cdot 08 + \cdot 05$.

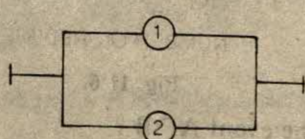


Fig. 11·5.

(ii) If D can work even if one of its components is functional, *e.g.*, a two-engine aeroplane or a bus with double tyres, then the reliability of D is

(*a*) $\cdot 92+\cdot 95$, (*b*) $\cdot 92+\cdot 95-\cdot 92\times\cdot 95$,
(*c*) $\cdot 92\times\cdot 95$, (*d*) $\cdot 92+\cdot 95+\cdot 92\times\cdot 95$.

REVIEW EXERCISE XI

1. If $S=\{S_1, S_2, \cdots, S_n\}$ is the sample space of a certain random experiment, then distinguish between S_1 and $\{S_i\}$, $1\leqslant i\leqslant n$.

2. Prove that the events A, B on the same sample space are complementary if and only if they are both (*i*) mutually exclusive and (*ii*) exhaustive.

3. Prove that

$$P(A\cap B)\leqslant P(A)\leqslant P(A\cup B).$$

4. The odds in favour of a certain event are as 1 is to m. What is the probability of its happening ?

5. Deduce the relation $P(\bar{A})=1-P(A)$ from the addition theorem of probabilities.

6. Interpret the adjoining diagram. Also, answer the following questions :

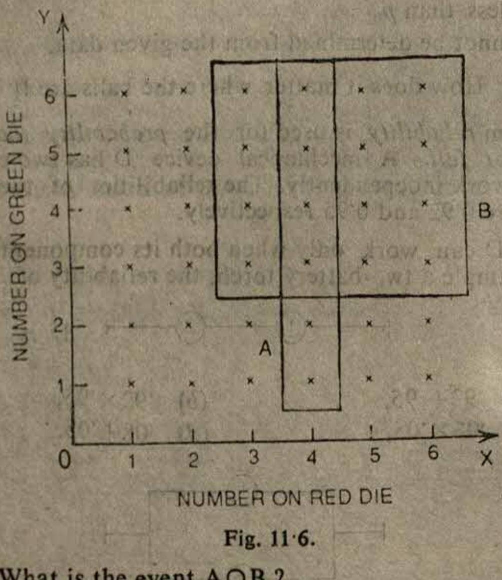


Fig. 11·6.

(*a*) What is the event $A\cap B$?

(*b*) What are P(A), P(B), $P(A\cap B)$ and $P(A\cup B)$?

Now verify the generalized addition theorem and the multiplication theorem of probabilities for the events A and B.

7. Verify the following law by using Venn diagrams :

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

In the light of the above law, comment on the following problem :

Among 10 friends, 8 are tinkers, 6 are drinkers and 2 are thinkers. What is the probability that one selected at random will be a tinker, a thinker and a drinker too ?

8. Are the following events independent or not ?

(*a*) A coin is tossed and a dice is rolled. A is the event that the coin shows head and B is the event that the dice shows a 6.

(*b*) From a bag containing 6 blue marbles and 4 red marbles, two marbles are drawn in succession. A is the event that the first marble is blue and B is the event that the second marble is red.

(*c*) In the random experiment of rolling a red die and a green die, A is the event that the red die shows an even number, B is the event that the green die shows an even number and C is the event that the sum of the numbers on the die is even.

9. A box contains 3 coins, two of which are fair but the third of which has a Head on both sides. A coin is selected at random and tossed twice. What is the probability of getting

(*a*) tail both times ? (*b*) head both times ?

(*c*) once a head and once a tail. Draw a tree diagram to help you. Can you answer without actually calculating this probability ?

10. At a party, there are five cigarette-smokers to each pipe-smoker. On an average, a pipe-smoker uses ten times as many matches as a cigarette smoker does. Assuming that no devices other than the matches were used in lighting the cigars and the cigarettes, what is the probability that a spent match, picked at random, after the party is over, was used by a cigarette-smoker ?

[**Hint.** Find the ratio of matches used by the two types of persons.]

11. A bag contains 100 slips numbered from 0 to 99. The number on the slips are replaced by the sum of their digits, and slips slipped back into the bag. A slip is drawn at random now.

What is the probability of the number on this slip being 0 (1, 2, 3,..., 18 respectively) ?

12. What is the probability of getting exactly 50 heads in 100 throws of a coin ?

[**Hint.** Do not jump to conclusions. The answer is far from half. Use the technique of Example 23. It would be quite interesting and instructive to calculate the probabilities of '1 head in 2 throws', '2 heads in 4 throws', '3 heads in 6 throws', '4 heads in 8 throws', '5 heads in 10 throws', and so on. Correct to 2 decimal palces, the respective values are 0·5, 0·38, 0·31, 0·27, 0·25 etc.]

13. Comment briefly on the following statements :

(*a*) $\{1, 2, 3, 4\}$ and $\{4, 1, 2, 3\}$ are sample spaces of different random experiments.

(*b*) Every random experiment has infinitely many sample spaces.

(*c*) On the sample space $\{a, b, c\}$ at the most 6 events, viz., $\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}$, and $\{b, c\}$ may be defined.

(*d*) Doctor (to the patient) : You are suffering from a serious disease. 90% of the patients suffering from this disease, die. But do not worry. I treated nine patients suffering from this disease and they have all died ; you are the tenth.

(*e*) The Chieftain of a big tribe, wishes to increase the percentage of boys in the tribe. For this reason, he dictates that if a girl is born, then the parents of the girl must not have any more children. Assuming the probability of any child (born to any parents) being a boy to be 1/2, will the wish of the Chieftain be fulfilled ? Will it increase the percentage of girls in the tribe ?

(*f*) Given that 90% of the cancer patients die, who has a greater probability of life, a cancer patient or a rich woman ?

14. In a family there are three children. What is the probability that the family has (*i*) no male child, (*ii*) at least one male child, assuming that the chances of a child being a male or a female are equal ?

(*Roorkee, 1984*)

15. In a simple throw with two dice, find the probability of having 8 or 11 or 12.

(*A.I.S.S.C.E., 1986*)

16. Three six-faced dice are tossed together. Find the probability of obtaining a sum of 15 of all the three numbers on the dice. (*Roorkee, 1983*)

17. In a carton containing 60 eggs, 48 are fresh and 12 are a day old. If two eggs are picked up one by one at random, what is the probability that

(*a*) both are fresh ?

(*b*) both are a day old ?

(*c*) the first is fresh and the second a day old ?

18. Consider an electric wire to which 12 bulbs are attached here and there. The bulbs on the wire light if and only if all of them are good (*i.e.*. not fused). If the bulbs have been selected at random from a box of 100 bulbs containing 10 fused bulbs, what is the probability that the bulbs on the wire will light ?

19. A manufacturer supplies plates in boxes of 100. A shopkeeper examines 5 plates from each box. If any of these is chipped, he rejects the box. In the other event, he buys the box. If a particular box contains 10 chipped plates, what is the probability that the shopkeeper will purchase it ?

20. Explain the fallacy in the so-called paradoxes given below :

(*a*) On a particular day, Mohan Bagan and East Bengal are playing a football match. Let A be the event that the Mohan Bagan wins and B the event that the East Bengal wins. The events are well-defined no matter against whom these teams are playing. But if they are playing against each other, A, B are mutually exclusive but not independent. In the other case, when they are not playing against each other, A, B are independent but not mutually exclusive. How can this happen ? Two events are either mutually exclusive or they are not ; they are either independent or they are not !

(*b*) An unbiased die is thrown 4 times in succession. The probability of not getting a six in any throw is $\left(\frac{5}{6}\right)^4$.

Hence the probability of getting a six at least once is

$$1-\left(\frac{5}{6}\right)^4=0\cdot518 \text{ approximately.}$$

On the other hand the probability of getting a six in one throw is 1/6. Hence the probability of getting at least one

six in four throws is $4(1/6)=2/3=0{\cdot}6$—much greater than $0{\cdot}518$. (In any case, since the probability is greater than $\frac{1}{2}$, it should be advantageous to bet on throwing a six at least once in 4 throws of a die.)

(c) The probability of getting a six on both dice in a single throw with a pair of dice is $\left(\frac{1}{6}\right)^2=\frac{1}{36}$. De Mere decided that it should be advantageous to bet on at least one six doublet (*i.e.* both dice showing a six) among 24 throws with a pair of dice $\left(24.\frac{1}{36}>\frac{1}{2}\right)$. It did not turn out to be true. Since the probability of getting a six on both dice in one throw is $\frac{1}{36}$ that of not getting a six on both dice in one throw is $\frac{35}{36}$. Since the successive throws a six doublet in any of the 24 throws is $\left(\frac{35}{36}\right)^{24}$. Hence the probability of getting at least one six doublet in 24 throws with a pair of dice is

$$1-\left(\frac{35}{36}\right)^{24}=0{\cdot}491 \text{ approximately},$$

$$<\frac{1}{2}.$$

21. Prove that the solution of the following problem depends on Fermat's last theorem :

A bag contain x white and y black balls. Another bag contains an equal number of balls but only z of them are black. Let A be the event that n ($\geqslant 3$) balls drawn from the first bag with replacement are either all white or all black. Let B be the event that the same number of balls drawn from the second bag, also with replacement, are black. What are possible values of n, x, y and z so that $P(A)=P(B)$?

22. A pair of dice is thrown n times. The probability of obtaining at least one double six is $1-\left(\frac{35}{36}\right)^n$. Refer to Ex. 20 above.

Find the least value of n, so that this may be more than $0{\cdot}5$. Can you now say for how many throws should De Mere have settled in order to win ?

[**Hint.** Use log tables or a calculator to calculate the probability for various values of n. Notice that since $\frac{35}{36} > 1$, $\left(\frac{35}{36}\right)^n$ goes on decreasing as n goes on increasing. What value of n would you start with in view of Ex. 20 above ?]

SUMMARY

The phenomenon *chance* can be assigned numerical values.

Sample Space : S={all possible distinct outcomes of a random experiment.}

Probability Function P on $S=\{S_1, S_2, \ldots\ldots, S_n\}$ satisfies the conditions

(*i*) $0 \leq P(S_i) \leq 1$, $S_i \in S$, and

(*ii*) $P(S_1)+(PS_2)+\ldots\ldots+P(S_n)=1$.

Equally Likely Outcomes :

$P(S_1)=P(S_2)=\ldots\ldots=P(S_n)=\frac{1}{n}$, where

$S=\{S_1, S_2, \ldots\ldots, S_n\}$.

Event : A subset of the sample space S.

Algebra of events :

$\sim A$ or A' or $\bar{A}$ $\Leftrightarrow$ not A

AB or $A \cap B$ $\Leftrightarrow$ both of A and B

$A+B$ or $A \cup B$ $\Leftrightarrow$ at least one of A and B.

Special events :

A and B are mutually exclusive $\Leftrightarrow A \cap B=\phi$

$\Leftrightarrow P(A \cap B)=0$.

A and B are exhaustive $\Leftrightarrow A \cup B=S$

$\Leftrightarrow P(A \cup B)=1$.

Probability of Events :

If $A=\{S_1, S_2, \ldots\ldots S_k\}$,

Then $P(A)=P(S_1)+P(S_2)+\ldots\ldots+P(S_k)$, and

$P(\phi)=0$, by definition.

Addition Theorems :

(*i*) $P(A \cup B)=P(A)+P(B)$ if A and B are mutually exclusive.

(*ii*) $P(A \cup B)=P(A)+P(B)-P(A \cap B)$, in general.

Conditional Probability :

$$P(B \mid A)=\frac{P(B\cap A)}{P(A)},\ P(A)\neq 0.$$

Multiplication Theorem :

$$P(A\cap B)=P(B)\,P(A \mid B)=P(A)\,P(B \mid A)$$

Independent Events :

$$\text{A and B are independent} \Leftrightarrow P(A \mid B)=P(A)$$

$$\Leftrightarrow P(B \mid A)=P(B)$$

$$\Leftrightarrow P(AB)=P(A)\,P(B).$$

HISTORICAL NOTE

You must have realized by now that the probability theory, the study of the chance phenomenon, is not only a subject with great fascination and intrinsic value of its own but that it is also a subject which has great application value in almost all fields of human enquiry. It is really surprising that despite the fact that the chance phenomena have been present in man's environment quite since the beginning, except for stray references in the 15th and the 16th century the study of probability theory did not start before half the 17th century had passed. We owe it to the French nobleman Chevalier de Mere, that the foundations of the subject were laid, though not by the nobleman himself. He made his fortune in gambling by betting on at least one six with four dice and lost it fast enough by betting on at least one six-doublet in 24 throws with a pair of dice. He was completely surprised by this unexpected happening and wrote to his mathematician friend Pascal (1623–1662) demanding an explanation of the seemingly contradiction between his reasoning and observation. Pascal solved this puzzle (and some others too proposed by De Mere again !) for De Mere and informed his friend Fermat (1601–1665) about these developments, thereby making him also interested in this. Between the two of them, they laid the foundation of this charming subject which nevertheless has established itself, since then, a discipline of great utility. In has numerous applications in physical, biological and social science today. Believe it or not, but there is a very recently established discipline known as Statistical Mechanics. People who founded this discipline became world famous.

The first published work on the subject, published in 1654, is "De Ratiociniis in Ludo Aleae"—*On Reasoning in Games of Dice*, written by Huygens. The concept of *mathematical expectation* has its own origin in this book. "Liber de Ludo Aleae",—*The Book on Games of Chance*, written by Cardan (1501–1576) appeared posthumously in 1663. The next landmark was "Ars Conjectandi" of

Jacob Bernoulli (1654—1705), also published posthumously by his nephew Nicholas Bernoulli in 1713. Ars Conjectandi—*The Art of Conjecture*—was followed by "The Doctrine of Chances" of de Moivre (1667—1754) in 1718 In 1812 Laplace (1749—1827) published his great works "Theorie Analytique des Probabilities" and "Essai Philosophique sur les Probabilities", of which the later was a popular exposition of the subject meant for the general educated public, and an English translation of which—*A Philosophical Essay on Probabilities*—was brought out by Dover in 1951.

Till 1780, the theme song of the subject was—how to adjust stakes so as to be sure of winning in the long run, but important developments took place in the works of Russian mathematicians Chebycheff (1821—1894) and A. Markov (1856—1922). In the 20th century, the probability theory has been axiomatized (*i.e.*, you talk of a probability function satisfying certain conditions to be called the probability axioms rather than define the probability empirically, (based on experience) or as the limiting case of relative frequency). The first systematic studies in this direction were made by E. Borel, H. Steinhaus, P. Levy and A. Kolmogorov. The first published work on these lines was a monograph by Kolmogorov, available in English as *Foundations of Theory of Probability*, Chelsea, 1950.

□ □

JAKOB BERNOULLI (1654-1705)

Also known as James, Jacques, Jacob and Johann, the Swiss mathematician Jakob Bernoulli was born, and died too, at Basel.

Bernoulli's *Ars Conjectandi*, is the first treatise on the theory of probability, Cardan's and Huygens' *Ludo Aleae's* being introductory booklets only. It contains the first adequate proof of the binomial theorem for positive integral powers.

Bernoulli was fascinated the most by curves and calculus. He was the first to point out that a function need not be derivable at a maximum or a minimum point. He found the equations of the curves: catenary (the form assumed by a uniform flexible chain hanging freely under gravity ; equivalently the locus of the focus of a parabola rolling on a straight line), tractrix (a curve whose form is supposed to be ideal for a bearing which supports a revolving shaft exerting a lot of thrust parallel to the axis of rotation), and isochrone (a curve along which a pendulum would swing in the same period no matter what the amplitude). Bernoulli was, however, most partial to logarithmic spiral. He discovered many of its interesting properties and left instructions to the effect that this curve be inscribed on his tombstone. It is to Bernoulli that we owe the adjective *integral* in Integral calculus.

Random Variables and Probability Distributions

12·1. DISCRETE RANDOM VARIABLES

Have you heard of a *dust devil* ? It is not a devil (even if devils exist !). It is really a moving column of desert sand caused by natural phenomena. We shall now introduce to you a mathematical analogue of a dust devil, viz., a *random variable* (caused by chance phenomena). Analogue, in as much as it is no more a *variable* than a dust devil is a *devil* ; though it has something to do with *random* just as a dust devil has something to do with *dust*.

The sample spaces of random experiments are sets whose elements are outcomes of the experiments. These elements and their descriptions or notations appear in so many varieties as there can be, depending upon our interest, expression and even whim. They could be numbers or letters or even symbols. Let us consider the random experiment of tossing a fair coin twice. A sample space that we can consider is [HH, HT, TH, TT].

Now the way of specifying the outcomes by HH, HT etc. is optional. We have listed them as HH, etc. because that is self-explanatory. We could have denoted the first of these outcomes by 1, the second by 2 etc. and thus have the sample space {1, 2, 3, 4}. On the other hand, depending upon our interest in the outcome of the experiment, we could have done this job of assigning numbers to the outcomes in a more meaningful way than the arbitrary manner used above. We could count the *number of heads* in the outcome, say, and replace the outcomes by numbers in the following more meaningful way :

$$\text{HH} \leftrightarrow 2,$$
$$\text{HT} \leftrightarrow 1,$$
$$\text{TH} \leftrightarrow 1,$$
$$\text{TT} \leftrightarrow 0.$$

The agency which *associated a numerical value to each outcome* (or transformed the sample space S to {2, 1, 0}) is roughly a random variable as the term is understood. Since this agency is the function f on S defined as

$$f(\text{HH})=2, f(\text{HT})=f(\text{TH})=1 \text{ and } f(\text{TT})=0,$$

our random variable is a certain function from S onto $\{0, 1, 2\}$ or into **R**—the set of real numbers. To this extent, it is not a *variable* as understood by you in your calculus or algebra course ; it is *random* to the extent that its values are determined by the sample points which happen to be the outcomes of a random experiment. Formally,

Definition (Random Variable). *A real-valued function, whose domain is the sample space of a random experiment, is called a* **random variable.**

Remarks 1. Clearly, one can define more than one random variable on a given sample space. See *e.g.* Illustration 3.

2. Random variables are also known as **chance** or **stochastic** variables. (*Stochastic* is the Greek word for *chance*).

3. Random variables may be also defined on a sample space consisting of numbers. For example see illustration 3.

4. Just as a sequence, which strictly speaking is a function on the set of natural numbers, is normally specified and identified by the values it takes, similarly we shall often commit the crime of identifying the random variable with its range or the values that it assumes. Thus in case of tossing a coin, instead of saying that the random variable X is the function which associates 1 with head and 0 with tail (so that strictly, $X=\{(H, 1), (T, 0)\}$), we may say X is the number of heads obtained (so that $X=0$ or 1).

5. It is usual to denote a random variable by X. If X is a random variable on S and $S_i \in S$, then $X(S_i)$ is called the value of X at S_i.

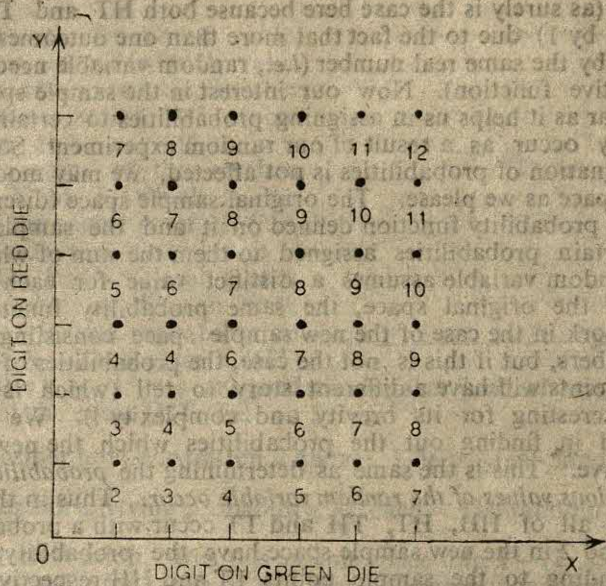
In this text, we shall be interested in random variables which take on only a finite number of values. Random variables defined on *discrete* sample spaces, automatically satisfy this condition. What should we call such random variables ? See the definition below to check on your guess.

Definition. *A random variable is said to be* **discrete,** *if its domain is a discrete sample space.*

Illustration 1. Let a single die be thrown. Let X have the value 1, if 4, 5 or 6 turn up, and the value 0 otherwise.

X is a random variable, with values 1 and 0. Is X discrete ?

Illustration 2. Let a red and a green die be thrown. Let X be the sum of the digits (number of dots) on the faces. Then X is a random variable, and its possible values are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.



The above figure shows the sample space (multitude of the dots) of the experiment. The numbers noted below the sample points (dots) indicate the values of the random variable. It is again an example of a discrete random variable.

Illustration 3. Let a single honest die be thrown and the number on the uppermost face be noted. Then S={1, 2, 3, 4, 5, 6}. The identity function on S is a random variable on S which assumes the values 1, 2,......, 6. Another random variable X on S may be defined as

X=0 if the number shown is six,

X=1 otherwise.

In this case X (1)=X (2)=X (3)=X (4)=X (5)=1 and X(6)=0 : X now assumes two values 0 and 1.

12·1·1. Probability Distributions

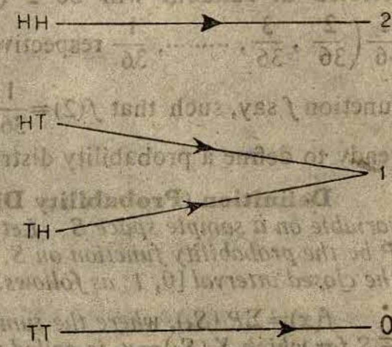
Observe that a random variable transforms a sample space S into a set T of real numbers. The sample space S here is the domain and the set T is the range of the random variable in question. This enables us to treat the outcomes of our random experiment as certain numbers and replace the original space by a new sample space consisting of numbers. For example, S={HH, HT, TH, TT} was replaced by {2, 1, 0}. Thus instead of regarding HH, HT etc. as the out-comes, we could consider 2, 1 etc. as the outcomes. Thus the new sample points are numbers instead of being symbols like HH etc. Also note that S and T do not always contain the same number of

elements (as surely is the case here because both HT and TH were replaced by 1) due to the fact that more than one outcomes may be replaced by the same real number (*i.e.*, random variable need not be an injective function). Now our interest in the sample space goes only as far as it helps us in assigning probabilities to certain events that may occur as a result of our random experiment. So long as this assignation of probabilities is not affected, we may modify our sample space as we please. The original sample space (discrete) has a certain probability function defined on it and the sample points have certain probabilities assigned to them the sum of which is 1. If the random variable assumes a distinct value for each sample point of the original space, the same probability function will clearly work in the case of the new sample space consisting of the real numbers, but if this is not the case, the probabilities of the new sample points will have a different story to tell (which is all the more interesting for its brevity and complexity !). We shall be interested in finding out the probabilities which the new sample points have. This is the same as determining the *probabilities with which various values of the random variable occur*. Thus in the above example, all of HH, HT, TH and TT occur with a probability $\frac{1}{4}$. Now 0 and 2 in the new sample space have the probability $\frac{1}{4}$ each corresponding to the sample points TT and HH respectively of S. However, 1 replaces the event {HT, TH} on S, the probability of which being the sum of the probabilities of HT and TH is $\frac{1}{2}$. Thus the probabilities of the various values 2, 1 and 0 of the random variable X have been determined as $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ respectively. This *distribution* of the total probability 1 to the various values of the random variable is known as *determining the probability distribution of the random variable X*. We may now forget about the original sample space S as we forgot the random experiment causing S once we had S. We may forget even the random variable X, if we remember the following table :

	0	1	2
	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

For all practical purposes, the above table, so far as this particular experiment is concerned, is quite enough.

Through the above table, we have assigned to each value of the random variable, another real number (called its probability) in this way—if x is a particular value of the random variable X on S, which is the image of the subset A of S (*i.e.*, if X (A)$=x$), then we assign to x, the number P (A). Recall that A being a subset of S is an event, and P (A) is the probability of this event A. Thus since X (HT) $=$X (TH)$=$1, so 1 was assigned the number $\frac{1}{2}$ which was the probability of the event {HT, TH} on S.



Consider now illustration 2 above. The original sample space has 36 elements, having probability $\frac{1}{36}$ each. The random variable X in that illustration, turns this sample space into T$=\{2, 3, \cdots\cdots, 12\}$ which is actually the range of X. Now 2 corresponds to one sample point *viz.* (1, 1) of S only. The probability of 2 is the same as that of (1, 1) in S and hence 2 is assigned the probability $\frac{1}{36}$ in the new sample space T. The points (1, 2), and (2, 1) of S give rise to 3 of T. Thus 3 is assigned $P(\{(1, 2), (2, 1)\})=P(1, 2)+P(2, 1)=\frac{2}{36}$ (the probability of the event {(1, 2), (2, 1)} of S). Similarly the point 4 of T arising from the event {(1, 3), (2, 2), (3, 1)} defined on S, has the probability $P(1, 3)+P(2, 2)+P(3, 1)=\frac{3}{36}$. Similarly for the other values in the table below :

Value of the random variable	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The second row lists the probabilities with which the various values of the random variable (listed in the first row) occur. In other words, the probability that a value of the random variable

selected at random will be 2 (3, 4, ..., 12 respectively) is $\frac{1}{36}\left(\frac{2}{36}, \frac{3}{36}, \ldots\ldots, \frac{1}{36} \text{ respectively}\right)$. Does it not give rise to a function f say, such that $f(2)=\frac{1}{36}, f(3)=\frac{2}{36}$ etc. ? We are now ready to define a probability distribution.

Definition (Probability Distribution). *Let X be a random variable on a sample space S. Let $T=X(S)$ be the range of X and let P be the probability function on S. Define a function f from T into the closed interval [0, 1] as follows. For each $x \in T$, let*

$f(x)=\Sigma P(S_i)$, where the summation is taken over all those $S_i \in S$ for which $X(S_i)=x$, is called the **probability distribution** (*or* **probability function** *or* **probability density function**) *of the random variable X.*

Remarks 1. For $x \in X, f(x)$ is the probability that X assumes the value x. Thus $f(x)=P(X=x)$. Hence $f(x)$ is the probability of an event defined on S. Accordingly, $0 \leqslant f(x) \leqslant 1$. Hence the range of f is a subset of [0, 1].

2. Once you have determined the probability distribution, it may be most of the times a good check (if not always !) to see whether or not the new probabilities add up to 1.

3. The probability distribution (or simply the distribution) of a random variable X is often given as a list of the distinct numerical values of X together with the associated probabilities. The list may be given horizontally or vertically. One can give a formula in place of the entire list.

Example 1. *Let E be the experiment of drawing a ball from an urn containing two red and one blue ball, and noting its colour.*

Solution. The sample space of the experiment is

$$S=\{r, b\},$$

and on the basis of balls being similar, except for colour, we may set

$$P(r)=\frac{2}{3}, \; P(b)=\frac{1}{3}.$$

Now let X represent the number of red balls drawn in the above experiment. Then X is a random variable with values 0 and 1. In fact, $X(b)=0$ (the case when a blue ball is drawn so that the number of red balls drawn is zero) and $X(r)=1$.

The probability function of X is defined by

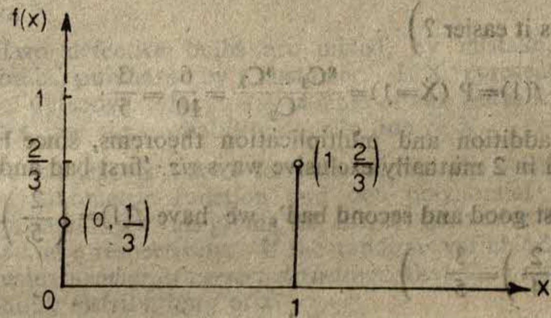
$$f(0)=P(X=0)=P(b)=\frac{1}{3}$$

$$f(1)=P(X=1)=P(r)=\frac{2}{3}$$

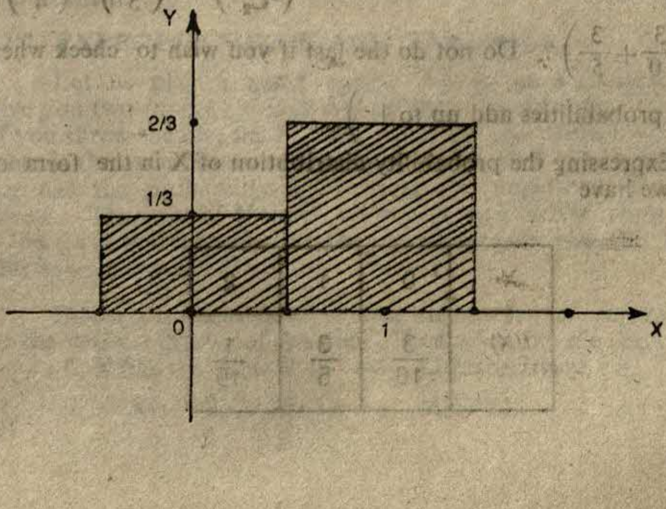
This may be represented in a tabular form as follows :

x	0	1
$f(x)$	$\frac{1}{3}$	$\frac{2}{3}$

Note that $\frac{1}{3}+\frac{2}{3}=1$. The graph below displays the same information more elegantly. By the way, what is the graph of the probability function here ? Is it the two bars or the two dots or both ?



We can also draw a probability histogram of the distribution. For this, the X-values are plotted on the x-axis and the $f(X)$-values



on the Y-axis. With each X-value as the centre, a vertical rectangle with area equal to the $f(X)$-value is drawn. The adjoining diagram shows the probability histogram of the distribution in the above example.

Example 2. *A housewife bought five eggs, among which two were bad. Find the probability distribution of the random variable X, which represents the number of bad eggs, if 2 eggs are taken at random.*

Solution. Let X be the number of bad eggs. Then X takes the values 0, 1, 2. If f is the probability distribution of the random variable X here, then

$$f(0)=P(X=0)=\frac{{}^3C_2}{{}^5C_2}=\frac{3}{10}$$

(Or, using multiplication theorem, $f(0)=\left(\frac{3}{5}\right)\times\left(\frac{2}{4}\right)=\frac{3}{10}$. Which way is it easier ?)

$$f(1)=P(X=1)=\frac{{}^2C_1 \cdot {}^3C_1}{{}^5C_2}=\frac{6}{10}=\frac{3}{5}.$$

(Or, using addition and multiplication theorems, since 1 bad egg could be had in 2 mutually exclusive ways *viz.* 'first bad and second good' or 'first good and second bad', we have $f(1)=\left(\frac{2}{5}\right)\times\left(\frac{3}{4}\right)+\left(\frac{3}{5}\right)\times\left(\frac{2}{4}\right)=\frac{3}{5}$.)

Similarly, $f(2)=P(X=2)=\frac{1}{10}$

(What did you calculate, $\left(\frac{1}{{}^5C_2}\right)$ or $\left(\frac{2}{5}\right)\times\left(\frac{1}{4}\right)$ or $1-\left(\frac{3}{10}+\frac{3}{5}\right)$? Do not do the last if you wish to check whether all the probabilities add up to 1.)

Expressing the probability distribution of X in the form of a table, we have

X	0	1	2
f(X)	$\frac{3}{10}$	$\frac{3}{5}$	$\frac{1}{10}$

Note that $\frac{3}{10}+\frac{3}{5}+\frac{1}{10}=1$.

(Graph the distribution yourself ! Also draw the probability histogram.)

EXERCISE 12 (*a*)

1. Let X represent the number of successes in a single throw of a die, where *success* means getting an even number. Express the probability distribution of the random variable X in the tabular form and sketch its graph.
2. Find the probability distribution f of the random variable Y, where Y represents the number of 'sixes' in two tosses of a die.
3. Two cards are drawn successively Y with replacement from a well shuffled pack of 52 cards. If X represents the number of aces drawn, find the probability distribution f of X.
4. Two defective bulbs are mixed, by mistake, in a lot of six bulbs, purchased by a customer. If X represents the number of defective bulbs in any sample of two bulbs from this lot of six, find the probability distribution of X.
5. A student is asked to write down the derivative of one trigonometric function and one polynomial function. The probability of his giving the correct derivative for the two is p and q respectively. If the random variable X assumes the value *number of correct derivatives*, then determine the probability distribution f of x.
6. From an ordinary deck of playing cards, cards are drawn one by one until either a diamond or a black card is drawn. Determine the probability distribution of the number of cards drawn.

12·2. EXPECTED VALUE AND VARIANCE

Let us play a game again. You throw a die and we are to give you two rupees for each dot on the uppermost face of the die. If you threw the die, say N times, what do you expect to gain and what is the expected average amount per game ? Each face of the die has the probability $\frac{1}{6}$ of being the uppermost face in every throw. Thus out of N times, we expect each face to show up N/6 times (in actual fact, the number will be roughly N/6 for large N.) Our sample space is {1, 2, ..., 6}.

Define a random variable X as the *number of rupees* you get on the various throws of the die. Thus when the die show up 1, 2, ..., 6, X has the value 2, 4, ..., 12 respectively, *i.e.*,

$$X(1)=2,\ X(2)=4,\ \ldots\ldots,\ X(6)=12.$$

Hence if f is the probability distribution of X, then

$$f(2)=P(X=2)=P(1)=\frac{1}{6},$$

$$f(4)=P(X=4)=P(2)=\frac{1}{6},$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots$$

$$f(12)=P(X=12)=P(6)=\frac{1}{6}.$$

Tabulating these values, we get

X	2	4	6	8	10	12
$f(X)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Now let us calculate the amount you expect to get when the die is thrown N times. Since we expect the die to show up a "1" $\frac{N}{6}$ times, these $\frac{N}{6}$ throws will bring you rupees $2\times\frac{N}{6}$. Similarly, the $\frac{N}{6}$ throws which result in a "2", would bring you Rs. $4\times\frac{N}{6}$ and so on.

Your total expected gain in rupees is

$$2.\frac{N}{6}+4.\frac{N}{6}+6.\frac{N}{6}+8.\frac{N}{6}+10.\frac{N}{6}+12.\frac{N}{6}$$

$$=N\left(2.\frac{1}{6}+4.\frac{1}{6}+6.\frac{1}{6}+8.\frac{1}{6}+10.\frac{1}{6}+12.\frac{1}{6}\right).$$

This is the amount expected in N throws. Therefore, the average amount expected per throw is

$$2.\frac{1}{6}+4.\frac{1}{6}+6.\frac{1}{6}+8.\frac{1}{6}+10.\frac{1}{6}+12.\frac{1}{6},$$

which is the sum of the various values of the random variable multiplied by the corresponding probabilities. This is an example of how the gamblers calculate their expected gains and make bets. As we told you in the previous chapter, determination of gains (or losses) was the motivating force for the development of the subject till towards the end of the previous century. The idea of mathematical expectation is nevertheless equally important in every field

where the subject finds application. We now make this idea precise.

Definition (Expected Value). *Let X be a finite random variable taking the values $x_1, x_2, \ldots\ldots, x_n$. For each $i=1, 2, \ldots\ldots, n$, let $p_i=p\,(X=x_i)$. Then the* **expected value** *of the random variable X (or the probability distribution of X), denoted by $E\,(\bar{X})$ or $\bar{\bar{X}}$, is defined as*

$$E\,(X)=p_1x_1+p_2x_2+\cdots+p_nx_n,$$

$$=\sum_{i=1}^{n} p_ix_i.$$

Remarks 1. The terms **mathematical expectation** or **expectation** or **mean** are also used for the expected value.

2. The formula for the mean of a probability distribution is similar to that for the mean of a frequency distribution, *viz.*, $\Sigma f_i \,.\, x_i/\Sigma f_i$. Here $\Sigma p_i=1$.

3. Like the mean of a frequency distribution, the mean of a probability distribution or of a random variable need not necessarily be a value of the variable. See Example 5.

4. If g is a function of X, the expected value of $g(x)$ is defined as

$$E\,(g(X))=\sum_i p_ig(x_i),$$

where as before, $p_i=P\,(X=x_i)$.

Example 3. *Compute $E\,(x)$ for the probability distribution*

x	1	2	3	4
$f(x)$	0.4	0.3	0.2	0.1

Solution. By definition,

$$E(x)=\sum_i x_ip_i,$$

$$=1\,(0{\cdot}4)+2\,(0{\cdot}3)+3\,(0{\cdot}2)+4\,(0{\cdot}1),$$
$$={\cdot}4+{\cdot}6+{\cdot}6+{\cdot}4$$
$$=2{\cdot}0. \qquad \textbf{Ans.}$$

Example 4. *The probability distribution of the random variable X, defined on the sample space {1, 2, 3} as X(n)=n−1 for =1, 2, 3, is given below.*

X	0	1	2
$f(X)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Find $E(X^2)$.

Solution. $E(X^2)=\sum_{i=1}^{3} p_i x_i^2$

$$=\frac{1}{4}\cdot 0+\frac{1}{2}\cdot 1^2+\frac{1}{4}\cdot 2^2$$

$$=\frac{3}{2} \quad \textbf{Ans.}$$

Example 5. *Ram and Shyam are playing with a die. They stake Re 1.00 each. The one who throws the higher number wins, and takes the total amount. If the number thrown by both is the same (i.e., in case of draw), they take back their stakes. Ram starts and throws a 2. What is the expectation at this stage ?*

Solution. If Ram wins, he gets Rs. 2.00. He invested Re. 1.00. Hence his gain is Re. 1.00. In case of a draw, he gets back his money and his gain in rupees is zero. In case Shyam wins, Ram suffers a loss of Re. 1. This can be expressed as a gain of −1 in rupees. Thus the values of the random variable in question, *viz. Ram's gain in rupees*, are 1, 0 and −1. Ram wins if Shyam throws 1 ; there is a draw if Shyam throws 2 and Ram loses if Shyam throws 3, 4, 5 or 6. The probabilities of these events are $\frac{1}{6}$, $\frac{1}{6}$ and $\frac{4}{6}$ respectively. Thus the probability distribution of Ram's gain is

X	1	0	−1
$f(X)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{3}$

Hence $\qquad E(X)=1\left(\frac{1}{6}\right)+0\left(\frac{1}{6}\right)+(-1)\left(\frac{2}{3}\right)=-\frac{1}{2}.$

Thus Ram expects to lose half a rupee. Note that this value is the *expected average* amount of loss or the *best estimate of the average amount of loss in the long run.*

Example 6. *A and B play a game with a coin. A stakes Re 1.00 and throws the coin 4 times. If he throws four heads, he gets his stake and also Rs. 3.00 from B. If he throws only three heads and they are consecutive, he gets his stake and also Rs. 2.00 from B. If he throws only 2 heads and they are consecutive, he gets his stake and also Re 1.00 from B. In all other cases B takes the stake money. Is the game fair ? (A game is said to be fair if the expectation of each player is zero.)*

Solution. Consider the sample space S={HHHH, HHHT, THHH, HHTT, THHT, TTHH,......others}. S has 2^4 elements. Let X be the random variable giving *A's gain in rupees.* If A throws 4 heads, we have the sample point HHHH and A gains Rs. 3. Thus X(HHHH)=3. Other values of X can be calculated similarly. The following table summarizes the situation:

Sample Point	HHHH	HHHT	THHH	HHTT	THHT	TTHH	Any other
Value of X	3	2	2	1	1	1	—1

Since out of a total of 16 sample points, exactly 1, 2, 3 and 10 produce the values 3, 2, 1 and —1 of X respectively, so the probability distribution of X is

X	3	2	1	—1
f(X)	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{5}{8}$

Hence $E(X)=3\cdot\frac{1}{16}+2\cdot\frac{1}{8}+1\cdot\frac{3}{16}-1\cdot\frac{5}{8}=0.$

Since A's gains and losses are B's losses and gains respectively, and A does not except to lose or gain anything, B's expectation is also 0. Hence the game is fair.

Example 7. *A travel agency is taking a one-week tour to a hill station. It offers the passengers insurance promising them Rs.*

15,000 00 in case of mishap during the trip. If the probability of a mishap is ·01, find a fair premium for this policy.

Solution. Let X be the random variable *payment (in rupees) per customer* by the company. Then the probability distribution of X is :

(Payment) X	(Probability) $f(X)$
0	·99
15,000	·01

Hence $$E(X)=0\times 0{\cdot}99+15000\times 0{\cdot}01,$$
$$=150.$$

Thus the company's expected payment per customer is Rs. 150.00. This may be viewed as a *fair* premium because on the basis of such a premium, the company *neither makes a profit nor does it lose anything in the long run.* In actual practice, the premium would be fixed at a higher rate making allowance for incidental expenses and profit by the company.

The expected value of a random variable gives us the best estimate of the average value of the random variable if the experiment is repeated a large number of times ; it tells nothing whatever about the way the various values are spread out. A measure of the dispersion of the various values is the *variance* of the random variable which we define as follows :

Definition (Variance). *Let X be a discrete random variable. Let $x_1, x_2, \ldots, x_n$ be the possible values af x, and $\bar{X}$ the mean of X. For each $i=1, 2, \ldots, n$, let $p_i=P(X=x_i)$. Then the positive number*

$$\sum_{i=1}^{n}(x_i-\bar{X})^2 p_i$$

is called the **variance** *of X and is denoted by Var (X) or σ_x^2.*

The number $\sqrt{Var\,(X)}$ is called the **standard deviation** *of X.*

Remark $\sigma_x^2=E[(X-\bar{X})^2]$.

Theorem. *If σ_x and $\bar{X}$ denote respectively the variance and mean of a random variable X, then*

$$\sigma_x^2=E(X^2)-\bar{X}^2.$$

Proof. $$\sigma_x^2=\sum_i p_i(x_i-\bar{X})^2,$$

$$=\sum_i p_i(x_i^2-2\bar{X}x_i+\bar{X}^2),$$

$$=\sum_i p_i\, x_i^2-2\sum_i \bar{X}p_ix_i+\sum_i p_i\bar{X}^2,$$

$$=E(X^2)-2\bar{X}\sum_i p_ix_i+\bar{X}^2\sum_i p_i,$$

$$=E(X^2)-2\bar{X}.\bar{X}+\bar{X}^2,\ \because \sum_i p_i=1,$$

$$=E(X^2)-\bar{X}^2.$$

The above formula may be more convenient in most of the cases for calculating the variance of a random variable.

Example 8. *Find the mean and the variance for the probability distribution*

X	-3	-1	0	4
$f(X)$	0.2	0.4	0.3	0.1

Solution. By definition,

$$\bar{X}=\sum_{i=1}^{4} x_ip_i,$$

$$=(-3)(0{\cdot}2)+(-1)(0{\cdot}4)+0(0{\cdot}3)+4(0{\cdot}1),$$

$$=-0{\cdot}6-0{\cdot}4+0+0{\cdot}4,$$

$$=-0{\cdot}6$$

Now the variance is

$$\sigma_x^2=\sum_{i=1}^{4}(x_i-\bar{X})^2\,p_i$$

Tabulating the values

x_i	p_i	$x_i - \bar{x}$	$(x_i - \bar{x})\, p_i$	$(x_i - \bar{x})^2\, p_i$
−3	0.2	−2.4	−.48	1.152
−3	0.4	−0.4	−.16	0.064
0	0.3	0.6	0.18	0.108
4	0.1	4.6	0.46	2.116
Total →	1.0	2.4	0.0	$3.440 = \sigma_{X^2}$

$$\therefore \qquad \sigma_{X^2} = 3{\cdot}44 \quad \textbf{Ans.}$$

Aliter As before, $\bar{X} = -0{\cdot}6$. Now,

x_i	x_i^2	p_i	$p_i x_i^2$
−3	9	0.2	1.8
−1	1	0.4	0.4
0	0	0.3	0.0
4	16	0.1	1.6
Total →			3.8

$$\therefore \qquad E(X^2) = \sum_i p_i x_i^2 = 3{\cdot}8$$

$$\therefore \qquad \sigma_{X^2} = E(X^2) - \bar{X}^2,$$

$$= 3{\cdot}8 - 0{\cdot}36,$$

$$= 3{\cdot}44 \quad \textbf{Ans.}$$

EXERCISE 12 (*b*)

1. A collection of eleven shirt pieces has four defective pieces. A customer buys three of them without inspection. Determine the probability distribution of the *number X of defective pieces* which the customer may get.

2. Calculate the expected value, the variance and the standard deviation for each of the following probability distributions :

(*a*)

X	0	1	2	3	4	5
$f(X)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$

(*b*)

X	-1	0	1	2
$f(X)$	.25	.25	.25	.25

(*c*)

X	0	1
$f(X)$	p	q

where $p+q=1$.

3. Calculate $E(X^2)$ for the parts (*b*) and (*c*) in question 1 above.
4. An honest die is thrown. What is the expected value of the number on the uppermost face ? What is its variance ? What is its standard deviation ?
5. By paying Re 1.00, a man gets to throw three honest dice. For each six he gets Rs. 2.00. What does he expect, to lose or to win ?
6. A man puts three coins in his pocket one by one deciding at random as to whether it is to be a 25 paisa coin or a 50 paisa coin. What is the expected totol value of the coins he puts in his pocket ?
7. The following table shows the profits of an ice cream vendor on certain types of days and the probabilities of such days :

Profit	*Type of day*	*Probability*
Rs 50	sunny	0·5
Rs 30	cloudy	0·3
Rs 5	rainy	0·2

What is his expected profit ? What is the variance of his profit ?

8. For the probability distribution given below, find k, E(X) and σ_x^2.

X	1	2	3	4	5	6
$f(X)$	k	$2k$	$3k$	$4k$	$5k$	$6k$

9. The roulette of a casino has 18 white, 18 red and 2 black slots. You can bet any amount on any colour. If the winning slot shows your colour, you get your stake plus an equal amount. Otherwise you lose your stake. Show that the expected profit of a hundred rupee bet on white is -5.26 in rupees. What does the negative sign show ?

[**Hint.** Your gain (in rupees) is 100 on a white and -100 on the other two colours. What are the probabilities of the various colours ? All bets in casinos are so fixed that the expected profit of players is always negative. That is why the casinos thrive.]

10. A botanist is asked which of the three botanical names are associated with three given plants. The botanist does not know which plant has which name, but lest he be considered ignorant, he assigns the names to the plant at random. Obtain the probability distribution of the *number of correct names*. Also find the expected number of correct matches.

[**Hint.** Which of the values from 0 to 3 is not assumed by the random variable in question ?]

12·3. BINOMIAL DISTRIBUTION

So far, we did not put any restriction on the number of elements in a sample space as long as they remained finite.

We shall now consider rather special random experiments whose sample spaces contain exactly two elements. Such experiments are abundant. As a matter of fact, whenever we are interested in *just one* particular outcome of the experiment, a sample space for this experiment containing just 2 elements can be considered. For example, in throwing a die, one of the six faces would be topmost. Suppose we were interested in getting a six (as in case of starting a game). Then we would have two outcomes—*six* and *not six* (others). We could denote them by 1 and 0 respectively. In selecting a balloon for a child, who likes red say, we could have a sample space {red, not red}. We could denote the *red* by 1 and *not red* by 0. Similarly in any other random experiment, if the interest lies in one single outcome, *this outcome*

could be denoted by a 1 and *not this outcome* by 0, giving us the 2-elementic sample space $\{1, 0\}$. Such a random experiment is called a **Binomial trial** (because of the two outcomes) or **Bernoullian trial** in honour of Jacob Bernoulli who studied series of such trials *under certain conditions* in his book *Ars Conjectendi* at length, leading him to the discovery of one of the most important and beautiful theorems in the theory of probability (which unfortunately is beyond the scope of this book !).

We shall now consider the conditions under which Bernoulli studied a series of trials. In the first place, since a Bernoullian trial is a random experiment, we may repeat it a number of times, thus getting a series of such trials. For example *we could toss a fair coin n times* (or toss n fair coins once) and note whether a head appears ; we could *draw a card at random from a deck n times* and see whether it is the Jack of diamonds. On the series of trials in question, we impose two conditions.

1. *Trials are independent.* As you know, this means that the probability of the desired event *viz.* the occurrence of the outcome of our interest is not affected by the results of the other trials in the series. Drawing a card and noting whether it is the Jack of diamonds will give us a series of independent trials if we replace the card drawn, shuffle the pack well and then draw the next card. However, if the drawings were being made without replacement, the corresponding series would not be independent.

2. *The probability of occurrence of the desired event is constant i.e., remains the same at each trial.* For example, if the card drawn from a pack is replaced and pack shuffled well, the probability of the Jack of diamonds remains 1/52 each time. That, in every series of independent trials, the probability of the occurrence of the desired event need not be the same follows from the example that follows. Let the trial consist of drawing a bead from a jar containing beads of several colours and noting whether the bead is white. We may suppose there are 20 jars containing 100 beads each but that the number of white beads in each differs from the others. We draw a bead successively from each jar. This series of 20 trials is independent because if a particular jar contains m white beads, the probability of a bead drawn from this jar being white ($m/100$) remains unaffected no matter what the colours of the beads drawn from the other jars are. Clearly, the probability of a white bead differs from jar to jar. One quite effective way of ensuring the condition of constant probability is to repeat the trial under identically the same conditions.

We shall be interested in series of independent Bernoullian trials in which the probability of the desired event remains constant throughout. It is customary to call the happening of the desired event a *success* and the other outcome as the *failure*. However, these words have little to do with their ordinary meaning. Suppose we

were subjecting a student to a series of tests of uniform difficulty and were noting whether he fails in the tests. Assuming that his ability to answer the questions in the tests is not affected by the other tests, a 'success' in the context of these Bernoullian trials would mean that he fails. We shall use the letters S and F respectively for the words success and failure.

As agreed, suppose we repeat a binomial trial n times. Let the constant probability of success at each trial be p. The probability of failure is then $q=1-p$ because F and S are mutually exclusive and exhaustive events. Each time we repeat the trial, we may have a success or we may have a failure. A very natural question that arises at this stage is :

(1) How many successes do we expect ?

Clearly, the minimum number of successes is 0. On the other hand we may have a success *every* time yielding n as the number of successes. This prompts us to introduce a random variable x, taking the $n+1$ values $0, 1, 2, \ldots\ldots n$ on the sample space of the experiment consisting of repeating these n trials and noting the number of successes. A better question than (1) above would be :

(2) What are the probabilities of the various values $0, 1, \ldots\ldots, n$, of the random variable *viz.*, *number of successes* ?

The following theorem answers this question.

Theorem. *In a series of n Bernoullian trials, with probability of success p in each trial, the probability that the number successes is exactly r is given by*

$$^nC_r p^r (1-p)^{n-r}.$$

Proof. In order to have r successes, in a series of n trials, we must have a success in r of the trials and a failure in the remaining $(n-r)$ trials. One scheme in which this happens, is :

$$\underbrace{SS\ldots\ldots\ldots S}_{r \text{ times}} \qquad \underbrace{FF\ldots\ldots\ldots F}_{n-r \text{ times}}$$

Since the trials are independent and probability of success is p at each trial, the probability of the scheme of successes and failure above is :

$$\underbrace{p.p\ldots\ldots\ldots p}_{(r \text{ times}} \qquad \underbrace{(1-p)\,(1-p)\ldots\ldots\ldots(1-p)}_{n-r \text{ times})}$$

$$=p^r (1-p)^{n-r}.$$

Obviously, there are other schemes in which r successes and $n-r$ failures may occur ; but the probability of each such scheme is

a product of n factors, r of which are p and $n-r$ of which are $1-p$ in some order. Because of the commutativity of multiplication of real numbers, this product can be written as $p^r(1-p)^{n-r}$. The probability of r successes is obtained by adding the probabilities of all these schemes. The crux of the matter is now to determine the number of these schemes and that is easy. We may think of n dots in a row ; each dot is to be replaced either by an S or by an F in such a manner that there are exactly r S's. This amounts to choosing r dots out of n (to be replaced by an S). The number of different ways in which we can do this is ${}^nC_r$. Hence there are ${}^nC_r$ schemes yielding exactly r successes. The probability of each scheme being $p^r(1-p)^{n-r}$, the required probability is ${}^nC_r p^r(1-p)^{n-r}$.

We may now determine the probability distribution of the random variable x, *viz.* the *number of successes* in a series of n independent binomial trials with a constant probability of success at each trial.

Definition. *Given a positive integer n, and a real number p, $0 < p < 1$, the function P defined by :*

$$P(x)={}^nC_x p^x(1-x)^{n-x},\ x=0, 1, 2,\ldots\ldots, n$$

is called the **Binomial Distribution** or **Binomial Probability Distribution**.

Remarks 1. The various probabilities $P(x)$ are the terms in the binomial expansion of $(q+p)^n$, where $q=1-p$. Hence the name.

2. n and p are called the *parameters* of the Binomial distribution because they completely determine the distribution and for different values of n and p, different Binomial distributions are obtained. A Binomial distribution with parameters n and p will be denoted by B (n, p).

3. It is usual to denote ${}^nC_x p^x(1-p)^{n-x}$ by the symbol $b(x; n, p)$. Do not be afraid of the symbol. Read it as the *probability of x successes in a series of n binomial trials with a constant probability p of success at each trial.* If you take 10 different sets of values of x, n and p, and repeat the above sentence for these values, you would start feeling quite easy with the symbol.

4. Since the Binomial distribution occurs frequently in practical situations, tables giving the values of $b(x; n, p)$ for values of p from 0·01 to 0·50 in intervals of ·01 and for $n=2, 3,\ldots\ldots, 49$ are available. For some of these values of p and n, $b(x; n, p)$ correct upto 4 decimal places, has been tabulated in Appendix I at the end of the book. For use of these tables, see examples 10 and 11. Tables giving numerical values of some binomial coefficients are given in Appendix II. Using these tables may save you of lengthy calculations in some cases. You may use slide rule, log tables or pocket calculators for values of $b(x: n, p)$ not listed in the tables.

Example 9. *A fair coin is tossed six times. What is the probability of obtaining four or more heads ?*

Solution. Tossing a fair coin is a Bernoulli trial, with $p=1/2$. There are six trials.

The Binomial probability distribution with parameters 6 and 1/2, is given by :

$$P(x)={}^6C_x(1/2)^x(1/2)^{6-x},$$
$$={}^6C_x(1/2)^6,$$
$$=\frac{1}{64}\;{}^6C_x.$$
$$x=0,\ 1,\ 2,\ \ldots\ldots,\ 5,\ 6.$$
$$\therefore\ P(x\geqslant 4)=P\ (4)+P\ (5)+P\ (6)$$
$$=\frac{1}{64}\left[{}^6C_4+{}^6C_5+{}^6C_6\right]$$
$$=\frac{1}{64}\cdot(15+6+1)$$
$$=\frac{11}{32}=0{\cdot}344 \text{ approximately.}$$

Example 10. *Prove that :*

$$b\ (x\ ;\ n,\ p)=b\ (n-x\ ;\ n,\ 1-p).$$

Solution. We are required to show that the probability of x successes in a series of n independent binomial trials with constant probability of success at a trial p is the same as that of $n-x$ successes in a similar series with probability of success as $1-p$ at a trial. By definition,

$$b\ (x\ ;\ n,\ p)={}^nC_x p^x(1-p)^{n-x},$$
$$={}^nC_{n-x}(1-p)^{n-x}\,p^{n-(n-x)},$$
$$=b\ (n-x\ ;\ n,\ 1-p).$$

This relation enables us to use the tables for values of $p>0{\cdot}5$ even though the tables do not list values of $p>0{\cdot}5$.

Example 11. (*The Tea-tasting Lady*). *A certain lady claims that in most of the cases, she can tell by tasting the tea, whether milk was put into the cup and then tea-water poured, or first the tea-water was poured into the cup and then milk added. To test her claim, an experiment is performed. 9 pairs of cups of tea are prepared, each pair containing one cup of the two types mentioned. She is made to taste the 9 pairs and required to classify them according to the two types. If she is correct in at least 8 pairs out of 9, her claim will be accepted. We assume that the classification is done under independent and identical condition so that binomial distribution applies. If the lady is really skilled in the art she claims, p—the probability of her correct prediction must be high. Let us assume $p=0{\cdot}9$. The probability that lady's claim is proved (i.e., she tells 8 or 9 pairs correctly), is equal to*

$$b(8\,;9,0{\cdot}9)+b(9\,;9,0{\cdot}9) \qquad (n=9,\ p=0{\cdot}9\,;\ x=8,\ 9 \text{ resp.})$$
$$=b(9-8\,;9,0{\cdot}1)+b(9-9\,;9,0{\cdot}1) \qquad (\textit{see example } 10)$$
$$=b(1\,;9,0{\cdot}1)+b(0\,;9,0{\cdot}1).$$

Solution. We shall now use tables in appendix I to calculate the above value. Take $b(1\,;\ 9,\ 0{\cdot}1)$. Here $x=1$, $n=9$, $p=0{\cdot}1$. Tables list the values of n and x vertically below n and x respectively. Locate $n=9$ in the first column. Against this 9, there are 10 values of x from 0 to 9 in the second column ; pick up $x=1$. We shall now look at the row of numbers corresponding to this $x=1$. Look at the values of p in the top row. They are ·01, ·05,......, ·50. Our interest is in $p=0{\cdot}10$. This is the third value in the top row. Consider the column headed by this 0·10 and see where it meets the row we have fixed earlier (corresponding to $n=9$ and $x=1$). The figure at the intersection is 0·3874. This is the required value Similarly,

$$b(0\,;9,0{\cdot}1)=0{\cdot}3874.$$

Hence the probability of the lady's claim being true is 0·7748.

On the other hand if the lady were just guessing, so that $p=0{\cdot}5$, the required probability is

$$b(8\,;9,0{\cdot}5)+b(9\,;9,0{\cdot}5)$$
$$=0{\cdot}0176+0{\cdot}0020 \qquad \text{(using tables)}$$
$$=0{\cdot}0196,$$

which is rather small.

12·3·1. The Mean and the Variance of Binomial Distribution

Theorem. *For Binomial probability distribution $B(n, p)$ given by*

$$P(X=x)={}^nC_xp^x(1-p)^{n-x},\ x=0,1,2,\ldots\ldots,n$$

(*i*) $E(X)=np$, (*ii*) $Var\,(X)=np(1-p)$.

Proof. (*i*) Let $q=1-p$. Then

$$p_x=P(X=x)={}^nC_xp^xq^{n-x}.$$

By definition,

$$E(X)=\sum_{x=0}^{n} xp_x,$$
$$=[0.{}^nC_0p^0q^{n-0}+{}^nC_1pq^{n-1}+2.{}^nC_2p^2q^{n-2}+\ldots\ldots$$
$$\ldots\ldots+r.{}^nC_rp^rq^{n-r}+\ldots\ldots+np^n],$$
$$=p\left[nq^{n-1}+\frac{n(n-1)}{1\,!}p^1q^{n-2}+\ldots\ldots\right.$$
$$\left.+\frac{n\,!}{(r-1)\,!\,(n-r)\,!}p^{r-1}q^{n-r}+\ldots\ldots+np^{n-1}\right],$$

$$=np\left[q^{n-1}+(n-1)p^1q^{n-2}+\frac{(n-1)(n-2)}{2!}\,p^2q^{n-3}\right.$$

$$+\ldots\ldots+\frac{(n-1)!}{(r-1)!\,(n-1)-(r-1)!}\,p^{r-1}q^{n-r}+\ldots\ldots$$

$$\left.+p^{n-1}\right],$$

$$=np(q+p)^{n-1},$$

$$=np, \text{ because } q+p=1$$

(*ii*) Now

$$\text{Var(X)}=\sum_{x=0}^{n} x^2p_x-(\text{E(X)})^2.$$

But $\displaystyle\sum_{x=0}^{n} x^2p_x=[0+1.{}^nC_1pq^{n-1}+2^2\,{}^nC_2p^2q^{n-2}+\ldots\ldots$

$$+r^2\,{}^nC_rp^rq^{n-r}+\ldots\ldots+n^2p^n],$$

$$=np\left[1.\,q^{n-1}+2(n-1)pq^{n-2}+3\,\frac{(n-1)(n-2)}{2!}\,p^2q^{n-3}\right.$$

$$\left.+\ldots\ldots+\frac{r(n-1)!}{(r-1)!\,(n-r)!}\,p^{r-1}q^{n-r}+\ldots\ldots+np^{n-1}\right],$$

$$=np\left[1.q^{n-1}+(n-1)pq^{n-2}+\frac{(n-1)(n-2)}{2!}\,p^2q^{n-3}+\right.$$

$$\left.\ldots\ldots+p^{n-1}\right]+np[(n-1)pq^{n-2}$$

$$+2\,\frac{(n-1)(n-2)}{2!}\,p^2q^{n-3}+\ldots\ldots$$

$$\left.+\frac{(r-1)(n-1)!}{(r-1)(n-r)!}\,p^{r-1}q^{n-r}+\ldots\ldots+(n-1)p^{n-1}\right],$$

$$=np(q+p)^{n-1}+np(n-1)p(q+p)^{n-2},$$

$$=np+n(n-1)p^2,$$

$$=np[1+(n-1)p]=np+n^2p^2-np^2.$$

$$\therefore \quad \text{Var (X)}=\sum_{x=0}^{n} x^2p_x-(\text{E}(x))^2,$$

$$=np+n^2p^2-np^2-n^2p^2,$$

$$=np(1-)p.$$

Remark. Var(X) may be written as npq, where $q=1-p$.

Corollary. *If the mean and variance of a binomial distribution with parameters n and p be μ and σ^2 respectively, then*

$$p=1-\frac{\sigma^2}{\mu}, \text{ and } n=\frac{\mu}{p}.$$

Example 12. *What are the mean and variance of a binomial probability distribution with parameters 16 and 0·5 ?*

Solution. Mean$=np=16\times{\cdot}5=8.$

Variance$=npq=8\times{\cdot}5=4.$

Example 13. *The mean and variance of a binomial distribution are 16 and 4 in some order. What is the probability of success at a given trial ?*

Solution. If the parameters of the given distribution are n and p, then we know that the mean is np and the variance is npq. Since $0 < p < 1$ (why) ? and $p+q=1$, therefore, $0 < q < 1$. This means that $npq < np$. Hence the variance is smaller than the mean. Therefore, the mean must be 16 and the variance 4. Hence

$$q=\frac{npq}{np}=\frac{4}{16}=0{\cdot}25.$$

$$\therefore \qquad p=0{\cdot}75.$$

EXERCISE 12 (*c*)

1. A coin is tossed six times. What is the probability of getting at least two heads ? (*A.I.S.S.C.E. 1984*)
2. A die is thrown six times. What is the probability of
   (*i*) no six ? (*ii*) more than four sixes ?
   (*D.B.S.S.C.E. 1984*)
3. Find the probability of having at least one tail in four throws with a coin. (*M.N.R. 1983*)

   [It is sufficient to compute just a single probability. Which one ? Two were sufficient for Exercise 1 above.]
4. The probability that an event A happens in one trial of an experiment is 0·4. Three independent trials of the experiment are performed. Find the probability that the event A happens at least once. (*I.I.T.J.E.E. 1980*)
5. What is the probability of throwing 4 double sixes in 400 throws of 2 dice ?
6. Determine the binomial distributions (*i.e.* find the probabilities of all values of the variable) whose parameters are given below :

   (*a*) $n=6$, $p=1/3$; use tables to write the probability of various terms.

   (*b*) $n=4, p=1/6$; leave the probabilities in fractions.

7. On the average, a marksman firing at a target, hits the bull's eye once in three shots. If he shoots six times, what are the chances that he hits the bull's eye

(*a*) twice ? (*a*) four times ? (*c*) not at all ?

8. Six coins are tossed 64 times. Show that the theoretical frequencies of 0 head, 1 head,....., 6 heads, are given by :

No. of heads	0	1	2	3	4	5	6
Frequency	1	6	15	20	15	6	1

[Theoretical frequencies are obtained by multiplying the probability with the total number of tosses.]

9. A pocket full of change is dumped on a table. If there are 22 coins altogether, find the probability of at least two heads showing.

[Would you calculate $P(2)+P(3)+\cdots\cdots+P(22)$? If yes, think hard ; your labours may be reduced rewardingly.]

10. If 3% of the electric bulbs manufactured by a company are defective, find the probability that in a sample of 100 bulbs (*i*) 0, (*ii*) 1, (*iii*) 2.
bulbs will be defective.

11. A coin is tossed n times. What is the chance that the head will present itself an odd number of times ? (*I.I.T.J.E.E. 1970*)

12. Find the binomial distribution whose mean is 3 and variance is 2.

13. The advertising cost of a new detergent powder is Rs. 500.00. It is believed that 10,000 persons will notice the advertisement and 10% of them will buy a bag each. Calculate the mean number of persons buying the detergent. If profit per bag of the detergent be Rs. 2.00, is advisable to advertise ?

[**Hint.** Compare the difference between expected gain, *i.e.*, profit, and advertisement cost.]

14. Sketch a graph of the binomial distribution for $n=4, p=1/4$. Calculate μ_x (mean) and σ_x (standard deviation).

15. A student calculated the mean of a binomial distribution to be 12 and its standard deviation as 4. Any comment ?

TEST YOUR UNDERSTANDING XII

In each of the following questions, tick the correct response :

1. f defined as follows

X	1	2	3	4	5
f (X)	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{4}$	0	$\frac{1}{2}$

(*a*) is a probability function on $\{1, 2, \ldots\ldots, 5\}$ because $\Sigma f(x) = 1$, (*b*) is not a probability function because one value of $f(x)$ is $-$ve, (*c*) is not a probability function because one value of $f(x)$ is 0, (*d*) is not a probability function because two values of $f(x)$ are equal.

2. The expectation of a player in a certain game is Rs. 5·00. This means that if he plays

(*a*) one game, he will win Rs. 5·00, (*b*) *n* games, he will win Rs. T *n*·00, $n=2, 3, 4, \ldots\ldots$, (*c*) a large number of games, he will win Rs. 5·00, (*d*) a large number of games, he will win Rs. 5·00 per game on an average.

3. For the binomial distribution B(100, 0·1),

(*a*) $E(X)=1$ and $\sigma x=3$,

(*b*) $E(X)=10$ and $\sigma x+9$,

(*c*) $E(X)=1$ and $\sigma x=9$,

(*d*) $E(X)=10$ and $\sigma x=3$.

4. The standard deviation of a binomial distribution is 3. Its mean can be :

(*a*) 4 but not 2, (*b*) 2 but not 4,

(*c*) 8 but not 10, (*d*) 10 but not 8.

5. If P is the probability function of $B(n, p)$ and $p+q=1$, then $P(x+1)$ is equal to $\lambda P(x)$ where λ is,

(*a*) $\dfrac{x+1}{n-x} \times \dfrac{p}{q}$,

(*b*) $\dfrac{x+1}{n-x} \times \dfrac{q}{p}$,

(*c*) $\dfrac{n-x}{x+1} \times \dfrac{p}{q}$,

(*d*) $\dfrac{n-x}{x+1} \times \dfrac{q}{p}$,

(Here is something to reduce your labours; once you have calculated P(1), $P(2)=P(1) X\lambda$; $P(3)=P(2) X\lambda$ and so on. But if you make a mistake somewhere, all the following terms will be

wrong too ; You had better check a term here and there by direct calculation.)

6. Let P be the probability function of $B(2M+1, \frac{1}{2})$. If $P(X \leqslant M)$ denotes the probability that X takes a value less than or equal to M etc., then :

(a) $P(X \leqslant M) < P(X > M) < \frac{1}{2}$,

(b) $P(X \leqslant M) = P(X > M) = \frac{1}{2}$,

(c) $P(X \leqslant M) \geqslant P(X > M) > \frac{1}{2}$,

(d) $P(X \leqslant M) < P(X < M) < \frac{1}{2}$.

[**Hint**. Various probabilities are the sums of certain terms in the expansion of $(\frac{1}{2}+\frac{1}{2})^{2m-1}$.]

7. 99 seeds are sown. Independently of others, the probability of each seed germinating is $\frac{1}{2}$. The probability that at least 75 seeds would germinate is (*a*) more than $\frac{1}{2}$ and equal to $\frac{3}{4}$, (*b*) less than $\frac{1}{2}$ and equal to $\frac{1}{4}$, (*c*) less than $\frac{1}{2}$ but not $\frac{1}{4}$, (*d*) more than $\frac{1}{2}$ but not $\frac{3}{4}$.

CHOR-SEPOY (चोर-सिपाही) : A game popular in rural India is played like this. Barring two players, the *Chor* and the *Sepoy*, others sit in a circle cupping their hands on a knee. The *Chor* has a small pebble in his hand and goes round the circle putting his hand in the cupped hands of the various players one by one. He leaves the pebble in somebody's hand. All the players close their hand after the *Chor's* visit. The sepoy watches from a distance but cannot see who has got the pebble. He makes a guess by pointing toward somebody at random. Let x denote the *number of correct guesses* when the game is played n times by 7 players in all. Suppose every time a correct guess is made, the *Sepoy* gets Rs. 2, and every time the guess goes wrong, he loses a rupee. Denote by Y the random variable *Sepoy's gain in rupees*.

8 The probability of r correct guesses is :

(a) ${}^nC_r(1/7)^r\,(6/7)^{n-r}$,

(b) ${}^7C_r(1/7)^r\,(6/7)^{n-r}$,

(c) ${}^7C_r(1/5)^r\,(4/5)^{5-r}$,

(d) ${}^nC_r\,(1/5)^r\,(4/5)^{n-r}$.

9. The probability distribution of x is :

(a) $B(5, 0{\cdot}2)$, (b) $B(n, 0{\cdot}5)$,

(c) $B(n, 0{\cdot}2)$, (d) $B(5; n, 1/5)$.

10. The set of values assumed by Y is :

(a) $\{1, 2, 3, 4, 5\}$, (b) $\{0, 1, 2, 3, 4, 5\}$,

(c) $\{1, 2\}$, (d) $\{-1, 2\}$.

11. For the game to be fair, instead of Rs. 2, the reward for a correct guess should be

(*a*) Re 1, (*b*) Rs. 3,
(*c*) Rs. 4, (*d*) Rs. 5.

REVIEW EXERCISE XII

1. Both of the bags A and B contain a pair of red and a pair of green socks. One sock is taken from each bag at random. The socks are interchanged before being put back into the respective bags. Determine the probability distribution of the number of red socks in bag A.

2. Five points are marked on the x-axis (any line for that matter) at intervals of 1 cm. Two points are chosen at random. Determine the probability distribution of the distance between the points. Show that the expected distance between the two points is 2 cm.

3. A card is drawn at random from a standard pack of 52 playing cards and the score noted. If ace has one point, picture cards 10 each, and other cards according to their denomination, then show that the mean and variance of the score are respectively : $6\frac{7}{13}$ and $9\frac{159}{169}$.

4. Calculate the expected value and variance of the following probability distribution (known as the **uniform** or the **rectangular** probability distribution) :

X	0	1	2	...	n
$f(X)$	$\frac{1}{n+1}$	$\frac{1}{n+1}$	$\frac{1}{n+1}$	...	$\frac{1}{n+1}$

5. A coin is tossed until a tail shows. Determine the probability distribution of X=the number of tosses. Is X a discrete random variable? Verify that the sum of all the probabilities of various values of X equals 1.

6. If X is a random variable and a is a fixed real number, then prove that :
(*a*) $E(aX)=aE(X)$. (*b*) $E(X+a)=E(X)+a$.
(*c*) $E(X-E(X))=0$.

7. Prove that if X is a random variable with mean μ and standard deviation σ, then the mean and variance of the random variable $(X-\mu)/\sigma$ are 0 and 1 respectively.

8. Would it be correct to apply binomial distribution to calculate the probability that it will rain for at least 15 days in July

next year if each day of July this year is treated as a trial and record kept of the rainy days (so that probability of a day in July being rainy may be calculated) ?

9. Assume that the probability that parents having black eyes will have a child with brown eyes is 0·25. Calculate the probability of at least 3 children having brown eyes if there are 6 children in a family, the parents having black eyes.

10. X is the random variable *number of failures preceding the first success* in a series of n Bernoullian trials with probability p for success. Show that :

$P(X=x)=(1-p)^x p,\ x=0, 1, 2, \cdots, n.$

11. The constant probability of success in a series of Bernoullian trials is 0·5. How many trials in the least should be performed so as to make the probability of at least one success more than 0·75 ?

12. **The Host.** Four friends wish to have a party, the host to which is to be decided by toss as follows. All the friends toss a fair coin simultaneously, more than once if need be, till one of them gets an outcome different from all the others, be it head or tail. This odd man out is to play the host. What is the probability that they would be able to decide the host in the first round of the tosses ? What will be this probability if there were n friends ?

13. **Random Walk.** A is waiting for B who is late. After a while, A gets restless and starts moving about along the footpath taking each step at random either in the Eastern direction or in the Western. What is the probability that after having taken 10 steps, he would be (*i*) back at the starting point ? (*ii*) exactly two steps away from where he started ? (More complicated problems of this nature form the hard core of the most modern studies regarding the motion of molecules in gases.)

(**Hint.** In order to finish at the starting point, how many steps must he move to the East and West in any order ? How many to the East and West in order to finish two steps to the right (left) of the original position ?)

14. Find out what is Galton's *quincunx* (board) and what does it have to do with Binomial distributions.

15. Verify that the sum of all the probabilities for a Binomial distribution turns out to be 1.

SUMMARY

Random Variable. A real-valued function whose domain is the sample space of a random experiment.

DISCRETE RANDOM VARIABLE

1. **Definition.** *A random variable with finite (or countable) range.*

2. **Probability Distribution.** (X is a random variable with domain S and range T). A function f from T into [0, 1], such that for all $x \in T$,

$$f(x)=\sum_{S_i \in S} p(S_i),$$

where $S_i \in S$ and $X(S_i)=x$.

— A function which assigns probabilities to all values of a random variable.

3. **Expected Value.** $\overline{X}=E(X)=\Sigma_i p_i x_i$

4. **Variance.** $\text{Var}(X)=\sigma_x^2=E(X-\overline{X})^2=\Sigma\, p_i(X_i-\overline{X})^2,$

$$=\Sigma_i p_i x_i^2-(\Sigma_i p_i x_i)^2,$$

$$=E(X^2)-\overline{X}^2$$

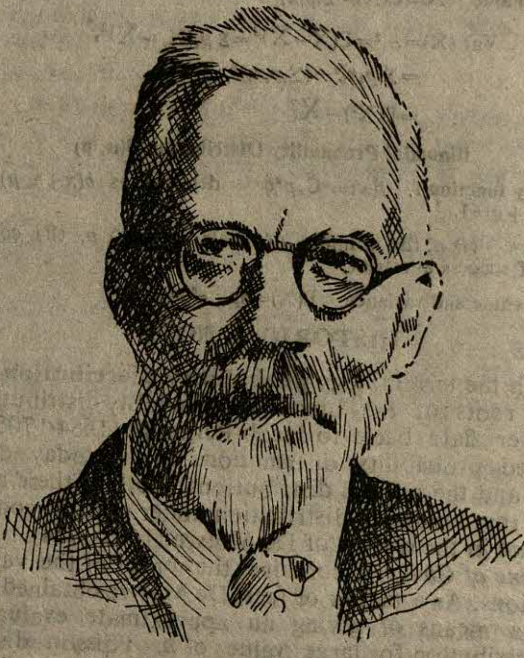
Binomial Probability Distribution B(n, p)

1. **Probability function f.** $f(x)={}^nC_x p^x q^{n-x}$ denoted as $b(x\,;n,p)$, $x=0, 1, 2, \ldots\ldots, n$, $p+q=1$.
2. **Parameters.** (*i*) n, the number of trials, and (*ii*) p, the constant probability of success at a trial.
3. **Expected value and variance.** $E(X)=np$; $\sigma^2_x=npq$.

HISTORICAL NOTE

Though the terminology **probability distribution** is not very old, yet the roots of the binomial probability distribution studied in this chapter date back to Jacob Bernoulli (1654-1705). Two of the most used probability distributions in day-today dealings are the Poisson and the normal distributions. Both of these are approximations of the binomial distribution under certain conditions and both originate in the works of De Moivre (1667-1754). His famous book *Doctrine of Chances* contains the proof of the validity of the approximations. As a matter of fact, De Moivre obtained the Normal function as a means of giving an approximate evaluation of the binomial distribution for large values of n. Poisson distribution is obtained from the binomial distribution by letting n become very large p very small but keeping np a fixed finite quantity. It should be noted that infinite and limiting processes were at that time in their crudest forms and the required theory was put on a sound basis later. This makes the efforts of these authors particularly worth the mention.

□ □



RONALD A. FISHER (1890-1962)

Sir Ronald A. Fisher was one of the most eminent scientists of the century who made important contributions to genetics and statistics. His work had a great impact on improvement of agricultural production all over the world. It demonstrates how statistics could be useful in real life.

CHAPTER 13

Correlation and Regression

13·1. INTRODUCTION

Do you have a friend who is both lazy and fat ? Do you have a sister who is both beautiful and intelligent ? Do you notice how your study hours change as the number of days to your approaching examinations decreases ? Do you think as the number of your days in class XII increases, so does the number of pages of your class-notes ? Do you think as the supply of a particular vegetable increases, its price falls ? Wouldn't it be nice if your parents' salary increased so that you could buy more clothes and more toys ? Such questions are of common occurrence in our daily life. Here, we are concerned with *two* (or more) characteristics of the item under observation, *e.g.*, laziness and weight of a friend ; brain and beauty of a sister ; number of your study hours and days left for your examination to begin, and so on. Not only that, consciously or otherwise, we also try to assess the pattern of change and possible relationships in the two characteristics. In this chapter, we shall consider such questions. We shall study objects only for *two* numerical characteristics like weight and height. Thus our data would consist of pairs of real numbers, giving us two variables. The object of this chapter is to learn techniques that would enable us to discover whether there exists some kind of relationship between the two variables, what is the strength of this relationship in case it is linear, and how to estimate the values of one variable from those of the other.

13·2. BIVARIATE FREQUENCY DISTRIBUTIONS

When we talked about things like mean basic salary of a group of employees, or the variance of marks of a collection of students, our data were concerned with *one* numeric characteristic of our units of observation. We shall now consider data which are concerned with *two* numeric characteristics.

Illustrations 1. Ages and number of fast friends of five girls are given below :

Name	*Age (years)*	*Number of friends*
Anupama	21	2
Nishi	20	3
Deepti	19	3
Parul	21	2
Shilpi	20	1

2. Doses of a medicine given and the time it takes in taking eeffect on three patients are given below :

Patient	*Dose (mg)*	*Time (sec)*
Jyoti	9	7
Asha	11	6
Kusum	10	8

Notice that each girl in the first illustration provides a pair of numbers ; one, when we record the age, and the other, when we record the number of friends. This clearly calls for two variables—*age* and *number of friends*. Similarly, in the second illustration, each patient provides two numbers, one corresponding to the *dose* and the other corresponding to the *reaction-time*. For this reason, data such as above are called **bivariate data**. A way, more meaningful than the above, of expressing these data is to use ordered pairs. This brings to fore the important fact that each unit of observation generates two numbers and is being observed with respect to two numeric variables. Thus the data regarding the five girls could be written as (21, 2), (20, 3). (19, 3), (21, 2) and (20, 1). Here the first co-ordinate gives the values of the variable 'age' and the second those of the variable 'number of fast friends'.

Suppose we have some bivariate data arising out of N units of observation. Let us denote the two variables by X and Y. Then our observations can be recorded as (X_1, Y_1), $(X_2\ Y_2)$, ..., (X_n, Y_n). The first co-ordinates refer to the variable X and the second to the variable Y. As in case of our data regarding a single variable, there is a reasonable chance of our being able to condense these data. Hopefully, not all the pairs (X_i, Y_i) are different. If a certain pair is repeated m times, we can call m the frequency of this pair. Listing only the distinct pairs together with their frequencies would condense the data somewhat. However, such a presentation would only serve the limited purpose which frequency distributions of one variable do. In fact, in so doing we miss the very aim bivariate data are meant to serve. The two variables in question may vary independently of each other in different ways We would like to classify our data in such a way as to be able to study such behaviour. Notice that the number of distinct values of X may be different from that of Y. For example, in the data consisting of the

two pairs (2, 4) and (2, 13), X takes only one distinct value *viz.* 2, but Y assumes two values *viz.* 4 and 13. Thus it would serve a better purpose if we classified our data collectively but also according to the two variables separately. We shall now explain how this two-way classification is effected.

Suppose among the N pairs (X_1, Y_1), (X_2, Y_2), $\cdots$, (X_N, Y_N), there are only m distinct values of X. Without loss of generality, we may denote them by $X_1, X_2 \cdots, X_m$. Similarly suppose that among the N y-coordinates only n are distinct. Denote these n distinct values by $Y_1, Y_2, \cdots, Y_n$. We shall now count how many times a value X_i occurs with a value Y_j. In other words, we shall count how many times the pair (X_i, Y_j), is encountered in the given N pairs. It is usual to denote the number of times it occurs by f_{ij}. Thus, we shall denote the frequency of the pair (X_i, Y_j) by f_{ij}. *The function which assigns the frequencies f_{ij}'s to the pairs (X_i, Y_j) is known as a* **bivariate frequency distribution.**

13·2·1. Two-way Frequency Tables

In case of univariate data, frequency tables involved two rows or two columns, one listing the variable-values and the other the corresponding frequencies. Now we have two variables. What shall we do ? One possible way could be to make two separate tables, one corresponding to each variable. However, that would be a very inadequate way. The two banks of a river do not form the river. A river is much more than its banks. So is the case with bivariate data. Bivariate data are much more than two sets of univariate data. Here, we are interested in discovering any inter-relations and inter-actions that exist between the two variables. Most of the scientific discoveries are made by discovering relationships that exist between two variables. Most of our decision-making in daily life is done on the basis of such relationships. Separate tables would contribute nothing towards this goal. To achieve this goal we represent bivariate data in a cross-table known as a **two-way frequency** (also **contingency/correlation**) table. The rows are used to represent one variable and the columns are used for the other. The various variable values are written in a row on top of the table for one variable, and in a column to the left of it for the other. The frequencies corresponding to a pair of values are written in the cell at the intersection of relevant row and column. For example, let X and Y denote respectively the marks of 50 boys in English and Mathematics, the maximum marks being 5 in each subject. Then each of X and Y varies from 0 to 5. These values are shown in Table 13·1 below.

The actual data observed are listed in the body of the table enclosed by bold lines. The entry 5 in the first cell of the second row (equivalently second cell of the first column) refers to the values X=0 and Y=1. All students getting 0 marks in English and 1 in Mathematics were put in this cell. There were five such students

Table 13·1 :
Marks of fifty students in English (X) and Mathematics (Y)

Y \ X	0	1	2	3	4	5	Totals
0	1	–	–	2	–	2	5
1	5	–	1	–	–	3	9
2	–	4	4	9	–	–	17
3	–	1	7	6	1	1	16
4	–	–	–	1	1	1	3
5	–	–	–	–	–	–	0
Totals	6	5	12	18	2	7	50

and so we had five tallies in this cell which we have replaced by the figure 5. 5 is thus the frequency of the pair (0, 1) because this pair of observations was obtained five times. The entry 9 in the table shows that the frequency of the pair (3, 2) is 9, or that 9 students obtained 3 marks in English and 2 in Mathematics. A dash indicates zero frequency. In general, the entry at the intersection of the column $X=X_i$ and the row $Y=Y_j$ is the frequency of the pair (X_i, Y_j) and is denoted by f_{ij}.

Remark. The last row in Table 13·1 is redundant and should be omitted, there being no students with 5 marks in Mathematics. The last column, however, has to be retained. Seven boys have obtained 5 marks in English. (And don't you pooh-pooh ! The English-test was on spellings and the Mathematics-test on tables of 17 to 21).

13·2·2. Marginal Distributions

You would notice that the Table 13·1 also shows row-totals and column-totals, which individually add up to 50. What do these totals represent ? Consider the total 6 of the first column obtained by summing the entries 1 and 5. It shows that there are 6 students who have obtained 0 marks in English, 2 of the six having obtained zero marks in Mathematics, and 5 having obtained 1 mark in Mathematics. To sum up, no matter what are their marks in Mathematics, 6 students have obtained zero marks in English.

Similarly, the column-total 5 shows that irrespective of their marks in Mathematics, 5 students have obtained 1 mark in English. This means that if we consider the column totals, we shall obtain the univariate frequency distribution of X, the marks obtained in English. Similarly, the row-totals provide the univariate frequency distribution of Y, the marks obtained in Mathematics. These univariate distributions obtained by considering the *marginal* totals are known as the **marginal frequency distributions** of X and Y associated with the bivariate frequency distribution given by Table 13·1 above.

Suppose that we have a bivariate frequency distribution of X and Y with X taking the *m* distinct values $X_1, X_2, \ldots, X_m$ and Y taking the *n* distinct values $Y_1, Y_2, \ldots\ldots, Y_n$ with f_{ij} as the frequency of the pair of values (X_i, Y_j). To obtain the marginal distribution of X, we shall hold the value of X fixed at X_i say, and sum up the frequencies of the pair (X_i, Y) letting Y range over all its values. Thus frequency of X_1 is the sum of the frequencies of the pairs $(X_1, Y_1), (X_1, Y_2), \ldots\ldots, (X_1, Y_n)$, which happens to be precisely the marginal total of the first column (See Table 13·2).

Table 13·2.

Y \ X	X_1	X_2	.....	X_i	.....	X_m	Totals f_Y
Y_1	f_{11}			f_{i1}			
Y_2	f_{12}			f_{i2}			
⋮	⋮			⋮			
Y_j	f_{1j}		---	f_{ij}	....	f_{mj}	$\sum_{i=1}^{m} f_{ij}$
⋮	⋮			⋮			
Y_n	f_{1n}			f_{in}			
Totals f_X	$\sum_{j=1}^{n} f_{1j}$			$\sum_{j=1}^{n} f_{ij}$			$\sum_{j=1}^{n} \sum_{i=1}^{m} f_{ij}$

Thus the frequency of X_1 is $f_{11}+f_{12}+\dots+f_{1n}$, or $\sum_{j=1}^{n} f_{1j}$. In general, the frequency of X_i is $\sum_{j=1}^{n} f_{ij}$, $i=1, 2, \dots, m$. The generic name given to these frequencies would be f_X.

Similarly, the frequency of Y_j (the sum of all the frequecies in the jth row) is $\sum_{i=1}^{n} f_{ij}$, $j=1, 2, \dots, n$. These frequencies are listed in the column 'Totals f_Y' above.

Remarks 1. $\sum_{j=1}^{n} f_{ij}$ denotes the number of observations for which $X=X_i$ no matter what their Y-value. $\sum_{i=1}^{m} f_{ij}$ is the number of observations having the Y-value equal to Y_j no matter what their X-value.

2. The sum of all the row-totals is equal to the total number of observations. Also, the column-totals add up to the total number of observations. Thus $\sum_{j=1}^{n}\left(\sum_{i=1}^{m} f_{ij}\right)=\sum_{i=1}^{m}\left(\sum_{j=1}^{n} f_{ij}\right)=N$, the total number of observations.

Example 1. *The marginal distributions of X and Y in the bivariate frequency distribution of table 13·1 are respectively the following :*

X_i (marks in English)	0	1	2	3	4	5
f_X (number of students)	6	5	12	18	2	7

Y_j (marks in Mathematics)	0	1	2	3	4	5
f_Y (number of students)	5	9	17	16	3	0

The first of these is effectively the frequency distribution of marks is English alone and the second that of marks in Mathematics alone.

13·2·3. Conditional Frequency Distributions

Let us once again consider the bivariate frequency distribution of Table 13·1. For a given value of Y, say Y=3, the frequencies with which the various X-values occur are given in the following table :

Y=3 X	0	1	2	3	4	5
f_X	0	1	7	6	1	1

Under the condition Y=3, the average or the mean value of X is, therefore,

$$\frac{\sum f_X X}{\sum f_X}=(1.1+7.2+6.3+1.4+1.5)/16$$

$$=42/16=2\cdot625.$$

Thus when Y=3, the average value $\bar{X}$ of X is 2·625. This gives us a pair ($\bar{X}$, Y) *viz.* (2·625, 3). Similarly, we can calculate the average value $\bar{X}$ of X corresponding to the other values of Y. These are listed below :

Y	0	1	2	3	4
$\bar{X}$ (approx. value)	3·2	1·9	2·3	2·6	4·0

The value 5 of Y being redundant, has been skipped.

The above distribution obtained from the given bivariate distribution is known as the *conditional distribution of X on Y*. The *conditioning variable* here is Y. There is of course another condi-

tional distribution (that of Y on X), associated with the given bivariate distribution in which X happens to be the conditioning variable. For each value of X, we find the average value of Y. For example, when X is 3, the relevant distribution to determine the average value $\bar{Y}$ of Y is

X = 3 Y	0	1	2	3	4	5
f_Y	2	0	9	6	1	0

Thus when $X=3$, $\bar{Y}=\dfrac{2.0+0.1+9.2+6.3+1.4+0.5}{2+0+9+6+1+0}$

$$=\frac{40}{18}=2{\cdot}2 \text{ approx.}$$

Similarly, calculating the other average values of Y, we get the following conditional distribution where X is the conditioning variable :

X	0	1	2	3	4	5
$\bar{Y}$ (approx. value)	0·8	2·2	2·5	2·2	3·5	1·4

Remark. We could have ignored the values with zero frequencies.

Example 2. *Determine the conditional distributions for the following bivariate frequency distribution where X is the measurement of a claw to the nearest mm and Y that of the body-length to the nearest mm of 555 crabs.*

Solution. Corresponding to the value $X=4$, there are two values 5 and 6 of Y having a non-zero frequency. Thus the average value of Y for $X=4$ is

$$\frac{1\times5+2\times6}{3}=\frac{17}{3}=5{\cdot}7.$$

Similarly, the average value of Y for $X=5$ is

$$\frac{7\times5+6\times6+5\times7}{18}=\frac{106}{18}=5{\cdot}9.$$

Table 13·3.

Claw-measurements and body lengths of a certain group of crabs

Y \ X	4	5	6	7	8	9	10	Totals
5	1	7	13	–	–	–	–	21
6	2	6	52	55	–	–	–	115
7	–	5	4	130	94	1	–	234
8	–	–	3	11	86	63	2	165
9	–	–	–	–	1	14	5	20
Totals	3	18	72	196	181	78	7	555

Calculating similarly the other average values of Y, we get the following conditional distribution of Y on X.

X	4	5	6	7	8	9	10
$\bar{Y}$ (approx. value)	5·7	5·9	6·0	6·8	7·5	8·2	8·7

Look at the above table carefully ! You cannot fail to see that *as the X-values increase so do the average Y-values.* This fact was not at all visible from the given table.

The other conditional distribution is given below :

Y	5	6	7	8	9
$\bar{X}$ (approx. value)	5·6	6·4	7·4	8·3	9·2

Again we find that as Y increases so does the average value of X. Do you think that on an average, the bigger the claw, the greater the body-length, and conversely too ?

Example 3. *Ten students obtained the following marks out of 10 in the first two tests in Physics :*

X (test 1)	7	6	5	8	5	8	7	7	8	7
Y (test 2)	7	6	7	3	6	1	2	0	1	2

Compute the relevant conditional distributions.

Solution. Here we have very limited data and no value-pairs are repeated. There is no need to make a table. When the X-score is 5, one Y-score is 7 and the other 6. The average Y-score for X=5 is, therefore, 6·5. For X=6, there is only one Y and this Y-value 6, is, therefore, also the average value of Y. Computing the other Y-values, we have

X	5	6	7	8
$\bar{Y}$	6·5	6·0	2·75	1·6

What is the trend ? As X increases, what happens to the average value of Y ?

The conditional distribution of X on Y is given below :

Y	0	1	2	3	6	7
$\bar{X}$	7	8	7	8	5·5	6

What do we notice now ? As Y increases, there seems to be a tendency for the average value of X to decrease, though not strictly. If we divide the data into lower Y-values (0, 1, 2 and 3) and higher Y-values (6 and 7), then we could say that the lower Y-group has higher $\bar{X}$-values and the higher Y-group has lower $\bar{X}$-values.

Remark. It would be a good idea to have a little practice in arranging a given bivariate data in the form of a 2×2 table. Try the data of Example 3. Take the given pairs (7, 7), (6, 6), (5, 7), ... one by one and go on putting a tally mark for each in the relevant cell of the required table. Compare with Table 13·4. Now count the tally marks in each cell and replace them with number (or dash in case there is no tally mark in a cell).

Table 13·4.

Scores in two tests of ten students

Y \ X	5	6	7	8	Totals
0			I		1
1				II	2
2			II		2
3				I	1
6	I	I			2
7	I		I		2
Totals	2	1	4	3	10

Have you been wondering why we are talking about marginal and conditional distributions associated with a bivariate distribution ? The purpose of the marginal distributions is obvious. In case we are interested in studying just one variable or the two variables separately with complete disregard to the other, then the marginal distributions are precisely what we need. However, we generally collect bivariate data wiih an interest in two characteristics *simultaneously*. To be precise, our interest lies in one or more of the following questions :

1. Are the two characteristics (or variables representing them) related ? If yes,
   (*a*) What kind of relation exists between them ? In particular, is it linear ? If yes,
   (*b*) Can the degree of this relation or association be determined quantitatively ?
2. Is it possible to know the values of one variable from a knowledge of those of the other ? If not, is it possible to have even a rough estimate for the values of one variable from those of the other ?

In the next section, we shall explore the answers to the first question. In other words, we shall first see how to examine whether there exists a *linear co-relation* between the two variables, and how

to measure its strength or degree if it does. The conditional distributions arise in our attempts to answer the second question.

In any statistical study, there is a population in the background which, more often than not, is somewhat vague. We wish to know more about it than we do. It may not be possible or desirable to study the whole of it. We, therefore, select a few items from the population at random without any bias. This limited collection of items from the population is called a *random sample*. From the study of this sample, we draw suitable conclusions about the whole population and also try to determine *quantitatively*, *how true* these conclusions might be.

It so happens that most of the time, we are studying *samples* rather than the whole population at large. Sometimes, both the variate values are random. For example, in a study of a school, we may select fifty students at random and record their weights and heights. Having ascertained that there does exist some relation between weights and heights, we may be interested in knowing what might be the weight of a student of a particular height. Now several students in the population would be having this particular height; but their weights would be *somewhat* different and not unique. The relationship between weights and heights can hardly be so perfect as a functional relationship of the type

$$\text{area of a circle} = \pi(\text{radius})^2,$$

where you are sure that given the radius r, the area *must* be πr^2. Here the same height may give us different weights. In fact, if somebody said that *all* students of a particular height have the *same* weight, we would laugh (and think such a person a nut or a screw loose in the brain !), yet we would believe that taller students would generally be heavier. Thus even though we agree there is a relation between *heights* and *weights*, *perfect mathematical relation* between the heights and weights would be to stretch the hypothesis of a *relationship* a bit too far. Wherever natural phenomena or human behaviour is concerned, the relationship between various characteristics are of the *height-weight type rather than radius-area type.* It may not be possible to predict the value *exactly*, but we may sort of *average* out. Given a particular height, we may consider the average weight of those students in our sample who have this height. Then we may say (and believe it too !) that a student of this height selected at random from this school would be expected to have *this average weight.* You can see now the role of conditional distributions in the question of determining the values of one variable on the basis of those of a co-related variable. We emphasize once again that the relationships here do not mean functional relations of the type $y=f(x)$ or $x=g(y)$, and that the values determined need not be the *exact* values; they are *estimated or expected values in the sense of average or representative values.* This is the technique of *regression*

(*estimation*) and forms the subject matter of the last section is this chapter.

EXERCISE 13 (*a*)

[Get all answers correct upto two places of decimal.]

1. Give five examples of bivariate data.
2. Tabulate the following bivariate data in the form of a bivariate frequency table :

   (1, 3), (4, 3), (2, 2), (3, 4), (4, 3), (2, 2), (1, 1), (3, 3), (3, 3), (4, 3), (1, 3) (1, 2), (1, 3), (2, 1), (3, 1).
3. Obtain the two marginal distributions of the data in Example 2.
4. The following table gives the scores X of 30 employees of a dress designing company in an aptitude test and the number of dresses Y designed by them in a week. Compute the two marginal distributions of X and Y. What type of data about the company do these distributions provide ?

Y \ X	5	6	7	8
3	1	4	–	–
4	2	3	2	2
5	–	2	5	9

5. Obtain the two conditional distributions associated with the bivariate frequency distribution of exercise

   (*a*) 2 above. (*b*) 4 above.
6. The following data give the number of stories (X) written by 15 authors over the number of months (Y) :

   (4, 2), (6, 6), (7, 7), (8, 8), (9, 8), (10, 12), (12, 10), (13, 11), (16, 12), (17, 11), (17. 15), (16, 16), (13, 12), (10, 10), (15, 11)

   (*a*) Arrange the data in a bivariate table.

   (*b*) Compute the two marginal distributions.

   (*c*) Obtain the two conditional distributions.

   (*d*) Plot the data points (preferably on graph paper) putting a dot for each point.

   (*e*) Do the dots show a pattern ? Can you draw a straight line close to most of the dots ? Is it the only such line ? If not, draw another. Can you say which is better in the

sense of being *as close to the dots as possible* ? What meaning do you assign to *close* here ?

7. Find the means X, Y, and standard deviations σ_X, σ_Y of the marginal distributions of Exercise 6 above. Consider the variable

$$x = \frac{X - \bar{X}}{\sigma x}$$

and
$$y = \frac{Y - \bar{Y}}{\sigma y}.$$

Plot the dots (x, y) for the values of (X, Y) in the above exercise. Does the shape of the pattern change ?

13·3. CORRELATION

We have seen that bivariate data may indicate some kind of relationship between the two variables, X and Y say, in question. This relationship may not be expressible by means of a mathematical formula. Yet there might be evidence of some rough pattern of related changes in the two variables. This relationship may be one of the following types :

1. *Cause and effect* type : Here change in the value of one variable causes a change in that of the other. For example, increase in the supply of a commodity may cause its price to fall.

2. *Common-cause* type : Here both variables change simultaneously due to some other agency. For example, both grey hair and weight of a person may increase because he is aging. Daily consumption of salads may increase with a corresponding decrease in that of bread because someone may be dieting. Variables like age and dieting here are known as *lurking* variables.

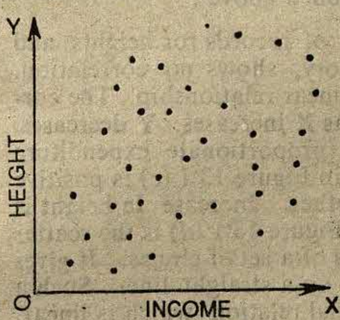
3. *Spurious* type : Sometimes a relationship may be there just by chance. For example the one in bad omens and mishaps ; or that between fish caught on a school day by a boy and the number of absentees in his class on that day.

It seems that the first stage in studying relationship between the two variables at hand is to use common sense and logical reasoning to determine whether a relationship could be spurious. Spurious relationships need not be bothered about. In the other event, if there is reason to believe that the two variables X and Y under consideration are *mutually related*, they are said to be **correlated.** When X and Y vary directly, *i.e.*, when they increase together (decrease together), the correlation is called **positive.** When X and Y vary inversely, *i.e.*, when an increase in the one is countered by a decrease in the other and conversely, the correlation is called **negative.** In the event of no relationship, the variables are said to be **uncorrelated.**

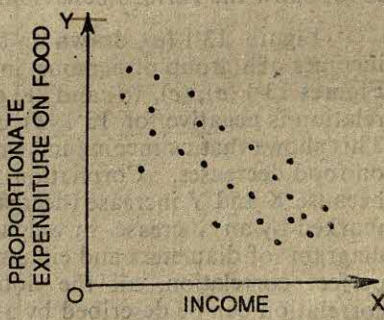
In what follows, *what* causes the correlation would not be our concern ; we shall be interested in determining *whether the correlation exists* and if it does, *what is the degree of this correlation.*

13·3·1. Scatter Diagram

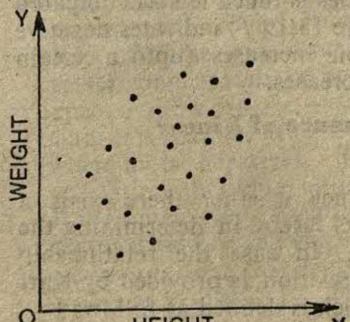
The question as to whether correlation exists, can be answered rather easily. Since our observations are ordered pairs of real numbers, they can be plotted as points (called dots and denoted as such) in the cartesian plane. This graph of the data is called the **scatter diagram** or the **correlation chart** of the data. It gives a visual impression of correlation, if there exists any. In case of correlation, it also gives an indication of the type of relationship that exists. To begin with, it makes easy the job of finding whether the correlation is positive or negative. Condensation of dots in the first and the third quadrants shows a positive correlation. Why ? Similarly, if most of the dots cluster in the second or fourth quadrant, the correlation is negative. The dots of the scatter diagram show a pattern in case of correlation. If the tendency of the dots is to cluster around a straight line, the relationship is called **linear.** If the dots lie on a line or a curve, the correlation is perfect. If the



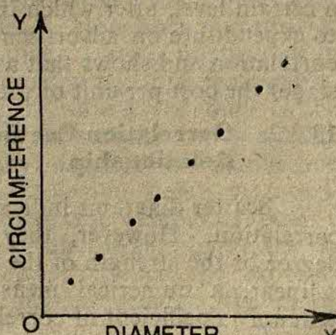
(*a*)



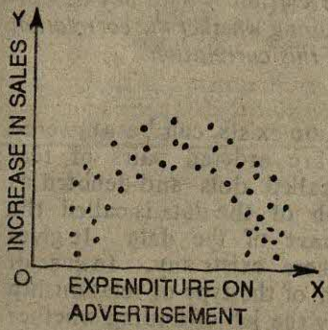
(*b*)



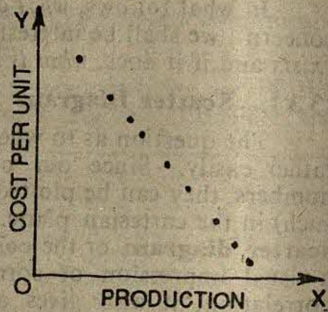
(*c*)



(*d*)



(e)



(f)

Fig. 13·1.

condensation of dots is around a smooth curve with a few waves, the relationship is called **curvilinear**. In case of no correlation, the dots scatter here and there without any pattern. Fig. 13·1 (*a*) to (*b*) show the various behaviours mentioned above :

Figure 13·1 (*a*), drawn on the basis of records for heights and incomes of a group of persons in a factory, shows no correlation. Figures 13·1 (*b*), (*c*), (*d*) and (*e*) show a linear relationship. The correlation is negative for 13·1 (*b*) because as X increases, Y decreases. This shows that as income increases, the proportionate expenditure on food decreases. Correlation shown in Figure 13·1 (*c*) is positive because X and Y increase (decrease) together. Increase in height is marked by an increase in weight too. Figure 13·1 (*d*) is the scatter diagram of diameters and circumferences of a set of circles. It gives perfect correlation, all the dots lying on a straight line. Such a correlation can be described by a functional relation which is linear. Figure 13·1 (*e*) depicts expenditure on advertisement and increase in sales. The sales increase with expenditure on advertisements upto a certain level, after which increase in sales is much less as compared to expenditure on advertisement. Figure 13·1 (*f*) indicates negative correlation and shows that as production increases (upto a certain stage) the cost per unit of production decreases.

13·3·2. Correlation Coefficient—Measure of Linear Relationship

Scatter diagrams indicate at a glance whether there exists a correlation. However, they are not as useful in determining the degree or the strength of the association. In case the relationship is linear, a numerical measure of correlation is provided by Karl Pearson's coefficient of correlation r, which is defined as follows :

$$r=\Sigma\left(\frac{1}{N}\cdot\frac{X-\bar{X}}{\sigma_X}\,\frac{Y-\bar{Y}}{\sigma_Y}\right)$$

Here $\bar{X}$, $\bar{Y}$ denote the mean values of X and Y respectively ; σ_X, σ_Y denote the standard deviations of X and Y respectively ; and N denotes the number of paired observations (X, Y).

If you are wondering why r above measures linear correlation, here is the explanation. Notice first of all that the relative position of dots remains unaltered no matter where you fix the origin. So we measure the two variables from their respective means ; this reduces our computation as the mean values of both $X-\bar{X}$ and $Y-\bar{Y}$ are zero. Division by σ_X, σ_Y makes the two variables independent of units of measurement. Thus X might be being measured in cm and Y in h, but $u=(X-\bar{X})/\sigma_X$ and $v=(Y-\bar{Y})/\sigma_Y$ are pure numbers. Algebraic manipulations involving both variables simultaneously become possible now. Hence the use of the standard units u and v. Next we use the key idea of relationship being linear. Suppose there is positive linear correlation as in Fig. 13·2. Notice that most of the dots would lie in the first and the third quadrant

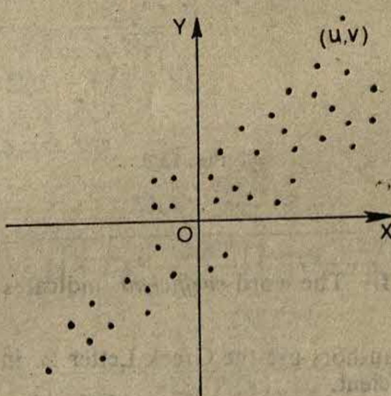


Fig. 13·2.

when *measured in standard units*. (Origin is $(\bar{X}, \bar{Y})$ and other values are scattered around it on both sides.) Now for a point (u, v) in either of these quadrants, uv is positive. Also, because of the greater condensation of dots in these quadrants, it is reasonable to expect u and v assume numerically bigger values for dots in these quadrants than for the dots in the second and the fourth quadrant (where uv is negative). Therefore, any measure of positive linear correlation should result in a positive Σuv, sum being extended to all the dots. Since by and large, the more the number of dots, the greater would be the value of Σuv, we divide the sum by N, the total number of dots, to neutralize this effect. Hence the expression $\Sigma uv/N$, or Pearson's r. For negative linear correlation, the majority of dots lie in the second and the fourth quadrant, each dot producing a negative uv term (See Fig. 13·3. In this case the dots in the second and fourth quadrants dominate those in the first and the third.

$\Sigma uv/N$ should be negative in this case. All said and done, Pearson's r is a very satisfactory measure of linear correlation.

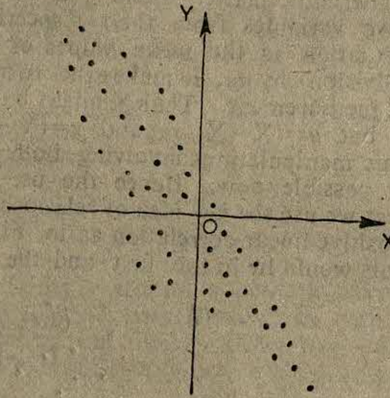


Fig. 13·3.

Remarks 1. The word *coefficient* indicates that r is a pure number.

2. Many authors use the Greek Letter ρ instead of r for the correlation coefficient.

3. As we shall show, $|r| \leqslant 1$ so that $-1 \leqslant r \leqslant 1$. The value $r=0$ indicates *no linear correlation*. $r=\pm 1$ indicate perfect linear relationship. If $r=1$, X and Y vary directly, correlation is positive, and all points of the scatter diagram lie on a straight line as in Fig. 13·1 (*d*). When $r=-1$, X and Y vary inversely ; correlation is negative and all points of the scatter diagram again lie on a straignt line as in Fig. 13·1 (*f*). The values of r between -1 and 0, and those between 0 and 1 indicate various degrees of coorelation. Higher (respectively lower) numerical values indicate high (respectively low) correlation. Thus *e.g.* the values $-·9$ and $·9$ of r are indicative of a very close association. A value like $-·2$ or $·2$ indicates very weak linear correlation. A positive sign of r indicates positive, and a negative sign shows negative correlation.

4. Since N, σ_X, and σ_Y are all positive, the sign of r is the same as that of $\Sigma(X-\bar{X})(Y-\bar{Y})$.

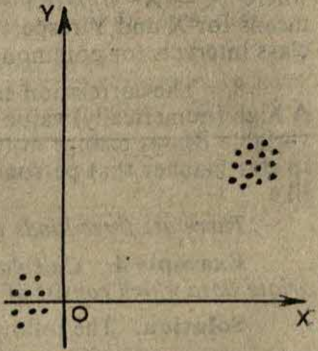


Fig. 13·4.

5. We emphasize once again that Pearson's *correlation coefficient measures linear correlation only.* Thus in cases like Fig 13·1 (*e*), r may turn out to be zero. Another situation where r is not a proper measure of association is, where the dots break into two distinct clusters like Fig 13·4.

6. The formula given above for r is generally not very convenient for actual calculations. For the purpose of calculations, the following variants of r may be more useful :

(*a*) $$r=\frac{S_{XY}}{\sqrt{S_{XX}}\ \sqrt{S_{YY}}},$$

where $S_{XY}=\Sigma(X-\bar{X})(Y-\bar{Y})$,

$S_{XX}=\Sigma(X-\bar{X})^2$,

and $S_{YY}=\Sigma(Y-\bar{Y})^2$.

(*b*) $$r=\frac{\Sigma XY-\dfrac{\Sigma X\Sigma Y}{N}}{\sqrt{\left[\left\{\Sigma X^2-\dfrac{(\Sigma X)^2}{N}\right\}\left\{\Sigma Y^2-\dfrac{(\Sigma Y)^2}{N}\right\}\right]}}$$

(*c*) $$r=\frac{\Sigma X'Y'-\dfrac{\Sigma X'\Sigma Y'}{N}}{\sqrt{\left[\left\{\Sigma X'^2-\dfrac{(\Sigma X')^2}{N}\right\}\left\{\Sigma Y'^2-\dfrac{(\Sigma Y')^2}{N}\right\}\right]}}$$

where $X'=X-a$, $Y'=Y-b$, a and b being suitable real numbers.

7. In case the values of the variables are repeated, we can modify the above formulae easily. Using f_x, f_y and f_{xy} for the frequencies of X, Y and (X, Y) respectively, we have the following formulae for r :

(*d*) $$r=\frac{\Sigma f_{xy}\ XY-\dfrac{\Sigma f_xX\ \Sigma f_yY}{N}}{\sqrt{\left[\left\{\Sigma f_xX^2-\dfrac{(\Sigma f_xX)^2}{N}\right\}\left\{\Sigma f_yY^2-\dfrac{(\Sigma f_yY)^2}{N}\right\}\right]}}$$

(*e*) $$r=\frac{\Sigma f_{xy}X'Y'-\dfrac{\Sigma f_xX'\Sigma f_yY'}{N}}{\sqrt{\left[\left\{\Sigma f_xX'^2-\dfrac{(\Sigma f_xX')^2}{N}\right\}\left\{\Sigma f_yY'^2-\dfrac{(\Sigma f_yY')^2}{N}\right\}\right]}}$$

where $X'=(X-a)/h$; $Y'=(Y-b)/k$; a and b being provisional means for X and Y respectively and h, k respectively are the equal class intervals (or common factors) for X and Y.

8. The correlation coefficient r should be used with caution. A high (numerically) value of r does not necessarily mean that one variable *causes* change in the other. It is because of novices using r in this manner that persons like Bengamin Disraeli make remarks like

There are three kinds of lies : lies, damned lies, and statistics.

Example 4. *Calculate the correlation coefficient for the bivariate data which consist of the pairs (3, 6), (0, 7), (4, 6) and (5, 1).*

Solution. The following table gives the desired computations.

X	$X-\bar{X}$	$(X-\bar{X})^2$	Y	$Y-\bar{Y}$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
3	0	0	6	1	1	0
0	–3	9	7	2	4	–6
4	1	1	6	1	1	1
5	2	4	1	–4	16	–8
Total 12		14	20		22	–13
$\bar{X}=3$		S_{XX}	$\bar{Y}=5$		S_{YY}	S_{XY}

Hence using formula (*a*) in remark 6,

$$r=\frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

$$=\frac{-13}{\sqrt{(14\times 22)}}$$

$$=\frac{-13}{2\sqrt{(11\times 7)}}$$

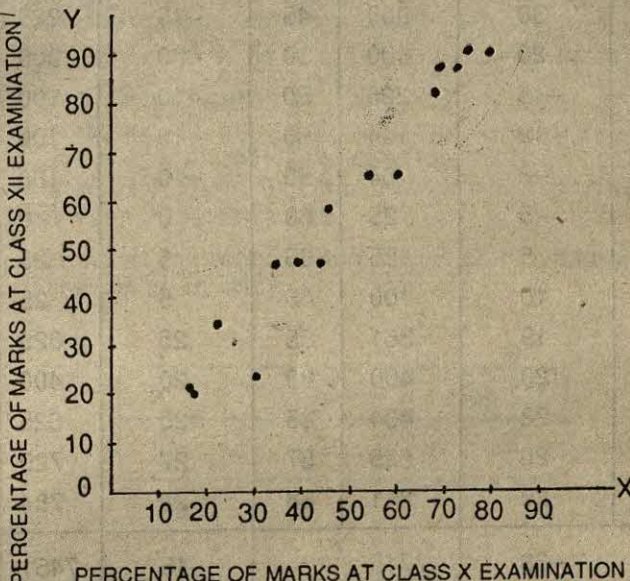
$$=-0{\cdot}74 \text{ approximately.}$$

Example 5. *The following table gives the marks (per cent) of 15 students at class X and class XII examinations. Draw a scatter*

X (Marks %)	20	21	30	40	45	48	52	55	65	70	79	80	82	85	89
XII (Marks %)	23	22	45	30	50	50	50	60	65	65	85	80	85	87	88

diagram to see if there exists a correlation between the marks at the two examinations and calculate Pearson's coefficient of correlation.

Solution. Let X and Y denote the marks at class X and XII examinations respecticely. Here N=15. The scatter diagram below shows that there is strong positive correlation. Thus it is meaningful to calculate Pearson's r in order to judge the strength of



the correlation. Measuring both X and Y from 60, we get the table as shown on next page. Using the data from that table.

We have $\Sigma X'=-39$, $\Sigma X'^2=7715$, $\Sigma Y'=-15$, $\Sigma Y'^2=7451$ and $\Sigma X'Y'=7349$. Also, N=15. Using formula (c) in remark 6,

$$r=\frac{\Sigma X'Y'-\dfrac{(\Sigma X')(\Sigma Y')}{N}}{\sqrt{\left[\left\{\Sigma X'^2-\dfrac{(\Sigma X')^2}{N}\right\}\left\{\Sigma Y'^2-\dfrac{(\Sigma Y')^2}{N}\right\}\right]}},$$

$$=\frac{7349-\dfrac{(-39)(-15)}{15}}{\sqrt{\left[\left\{7715-\dfrac{(-39)^2}{15}\right\}\left\{7451-\dfrac{(-15)^2}{15}\right\}\right]}},$$

$$=\frac{7349\times15-39\times15}{\sqrt{[\{7715\times15-1521\}\{7451\times15-225\}]}},$$

$$=0{\cdot}9715$$

$$=0{\cdot}97 \text{ approximately.}$$

X	X′ (= X − 60)	X'^2	Y	Y′ (= Y − 60)	Y'^2	X′Y′
20	−40	1600	23	−37	1369	1480
21	−39	1521	22	−38	1444	1482
30	−30	900	45	−15	225	450
40	−20	400	30	−30	900	600
45	−15	225	50	−10	100	150
48	−12	144	50	−10	100	120
52	−8	64	50	−10	100	80
55	−5	25	60	0	0	0
65	5	25	65	5	25	25
70	10	100	65	5	25	50
79	19	361	85	25	625	475
80	20	400	80	20	400	400
82	22	484	85	25	625	550
85	25	625	87	27	729	675
89	29	841	88	28	784	812
Total	−39 $\Sigma X'$	7715 $\Sigma X'^2$		−15 $\Sigma Y'$	7451 $\Sigma Y'^2$	7349 $\Sigma X'Y'$

13·3·3. Coefficient of Correlation for Continuous Variables

When the two variables being studied for correlation are continuous, then the data, quite often, are given in the form of a 2×2

WEIGHT (in Kg) \ HEIGHT (in cm)	90-100	100-110	TOTAL
25-30	25	19	44
30-35	17	9	26
Total	42	28	70

contingency (or *correlation*) *table* rather than as pairs of real numbers. For the purpose of illustration, suppose X and Y denote the heights and weights of a group of 70 students. The data are given as shown on page 718.

You are already familiar with frequency tables of bivariate discrete data. The only difference here is that cell-frequencies refer to whole classes rather than discrete values. Thus 25 in the above table relates to the class 90—100 so far as heights (in cm) are concerned and to the class 25—30 (in kg) so far as the weights are concerned. This means there are 25 students whose heights vary from 90 cm to 100 cm, with their weights varying from 25 kg to 30 kg. The meanings of marginal totals can similarly be modified. Thus *e.g.*, the marginal total 42 indicates that there are 42 children whose heights vary from 90 cm to 100 cm whatever their weights (25 kg to 30 kg or 30 kg to 35 kg).

As before, we can compute the marginal distributions of X and Y. These are given below :

X	90-100	100-110
f_X	42	28

Y	25-30	30-35
f_Y	44	26

Replacing the class-intervals by the respective mid-values, the distributions reduce to discrete frequency distributions, and you know how to deal with them. For purpose of illustration, let us consider the data of 70 students given above. Replacing the class intervals by their mid-values, the frequency distributions of X and Y reduce to the following :

X	95	105
f_X	42	28

Y	27.5	32.5
f_Y	44	26

The contingency table may be arranged as

Y \ X		90-100	100-110	f_Y
	MID VALUE \ MID VALUE	95	105	
25-30	27.5	25	19	44
30-35	32.5	17	9	26
f_X		42	28	70

25 is the number of students whose X-value is 95 and whose Y-value is 27·5. In other words, the frequency of the pair (95, 27·5) is 25. We can similarly consider the frequencies of other pairs of values of X and Y. The other f_{xy}'s are 19, 17 and 9. The correlation coefficient can be calculated by using any one of the formulae given in remark 7 earlier.

Example 6 *Calculate the coefficient of correlation between the marks obtained by 80 students in the terminal (X) and annual (Y) examination in a certain course, if the following table provides the data for the marks obtained.*

Y \ X	21-30	31-40	41-50	51-60	61-70	71-80	81-90	Totals
21-30	4	–	–	–	–	–	–	4
31-40	5	3	11	–	–	–	–	19
41-50	–	2	10	8	–	–	–	20
51-60	–	6	10	5	–	–	–	21
61-70	–	–	–	4	3	–	–	7
71-80	–	–	–	–	2	2	1	5
81-90	–	–	–	–	–	3	1	4
Totals	9	11	31	17	5	5	2	80

Solution. Let

$$X' = (X - 55{\cdot}5)/10, \quad Y' = (Y - 55{\cdot}5)/10.$$

		X	21-30	31-10	41-50	51-60	61-70	71-80	81-90				
		MID-VALUE	25.5	355	45.5	55.5	65.5	75.5	85.5				
Y	MID-VALUE	Y' \ X'	—3	—2	—1	0	1	2	3	f_Y	$f_Y Y'$	$f_Y Y'^2$	$f_{XY} X'Y'$
21-30	25.5	—3	4 —3(—3)4=36	–	–	–	–	–	–	4	—12	36	36
31-40	35.5	—2	5 —3(—2)5=30	3 —2(—2)3=12	11 —1(—2)11=22	–	–	–	–	19	—38	76	64
41-50	45.5	—1	–	2 —2(—1)2=4	10 —1(—1)10=10	8 0(—1)8=0	–	–	–	20	—20	20	14
51-60	55.5	0	–	—2(0)6	10 —1(0)10=0	5 0(0)5=0	–	–	–	21	0	0	0
61-70	65.5	1	–	–	–	4 0(1)4=0	3 1(1)3=3	–	–	7	7	7	3
71-80	75.5	2	–	–	–	–	2 1(2)2=4	2 2(2)2=8	1 3(2)1=6	5	10	20	18
81-90	85.5	3	–	–	–	–	–	3 2(3)3=18	1 3(3)1=9	4	12	36	27
		f_X	9	11	31	17	5	5	2	80	—41	195	162
		$f_X X'$	—27	—22	—31	0	5	10	6	—59			
		$f_X X'^2$	81	44	31	0	5	20	18	199			
		$f_{XY} X'Y'$	66	16	32	0	7	26	15	162			

The various calculations needed are arranged as in the table (page 721). The top entries in the various cells of the table are the frequencies f_{xy}. Thus 4 is the frequency corresponding to the values $X'=-3$, $Y'=-3$. The contribution $(-3).(-3).4$ to the sum $\Sigma f_{xy} X'Y'$ by this cell containing 4 is written just below 4 in this cell. The column headed by f_y gives the row-totals of the cell frequencies corresponding to a fixed value of Y′. Thus the second value 19 in this column is the sum of the frequencies in the second row which correspond to the fixed value -2 of Y′ and various values of X′. The column $f_y Y'$ lists the Products of f_y and Y′, *e.g.*, since 4 in the f_y —column is the frequency corresponding to the value -3 of Y′, the first entry in the column $f_y Y'$ is $4(-3)=-12$. The column $f_y Y'^2$ may now be obtained as the product $(f_y Y')(Y')$ etc. The column $f_{xy}X'Y'$ beside the $f_y Y'^2$ column lists the sum of the $f_{xy}X'Y'$—entries in various rows. Thus the second entry 64 in this column is the sum of the entries 30, 12 and 22 in the cells of the second row. The rows $f_x, f_x X', f_x X'^2$ and $f_x X'Y'$ may be interpreted similarly. It is not necessary to compute *both* of the $f_{xy}X'Y'$ —column and the $f_{xy}X'Y'$ —row. But it provides a good check, for though the various entries in the $f_{xy}X'Y'$ —row and the $f_{xy}X'Y'$ —column are different, they total up to the same figure. Now using the formula, we have

$$r=\frac{\Sigma f_{xy} Y'-(\Sigma f_x X'\Sigma f_y Y')/N}{[\{\Sigma f_x X'\,2-(\Sigma f_x X')^2/N\}\{\Sigma f_y Y'\,2-(\Sigma f_y Y')2/N\}]^{\frac{1}{2}}},$$

$$=\frac{162-(-59)(-41)/80}{[\{199-(-59)^2/80\}\{195-(-41)^2/80\}]^{\frac{1}{2}}},$$

$$=\frac{162-30{\cdot}2375}{[(155{\cdot}4875)(173{\cdot}9875)]^{\frac{1}{2}}},$$

$$=131{\cdot}7625/[2705\text{[illegible]}]^{\frac{1}{2}}$$

$$=131{\cdot}7625/164{\cdot}4776,$$

$$=0{\cdot}801.$$

Thus there is a strong positive linear correlation between the performance at the terminal and the annual examinations. Those who fared well at the terminal, also fared well at the annual examinations and conversely.

Example 7. *Show that Pearson's r lies betbeen -1 and 1. Also show that when $|r|=1$, then all the dots lie on a straight line.*

Solution. Assuming that our data consist of N pairs (X_i, Y_i), let us use the following formula for r :

$$r=\frac{\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\left[\sum_{i=1}^{N}(X_i-\bar{X})^2\sum_{i=1}^{N}(Y_i-\bar{Y})^2\right]}}.$$

Writing U_i for X_i-X and V_i for Y_i-Y, we get

$$r=\frac{\sum_{i=1}^{N}U_iV_i}{\sqrt{\left[\sum_{i=1}^{N}U_i^2\sum_{i=1}^{N}V_i^2\right]}}$$

Now *Schwarz inequality* for real numbers says that *if* $\{a_1, a_2, \ldots\ldots, a_n\}$ and $\{b_1, b_2, \ldots\ldots, b_n\}$ are two n-tubls of real numbers, then

$$\left(\sum_{i=1}^{n}a_ib_i\right)^2\leqslant\left(\sum_{i=1}^{n}a_i^2\right)\left(\sum_{i=1}^{n}b_i^2\right),$$

where equality holds if and only if for some real number k, we have

$$b_i=ka_i,\ i=1, 2, \ldots\ldots, n.$$

Letting n=N, $a_i=U_i$, $b_i=V_i$ in Schwarz inequality, we [illegible] that

$$\left(\sum_{i=1}^{N}U_iV_i\right)^2\leqslant\left(\sum_{i=1}^{N}U_i^2\right)\left(\sum_{i=1}^{N}V_i^2\right), \qquad \ldots(1)$$

and
$$\left(\sum_{i=1}^{N}U_iV_i\right)^2=\left(\sum_{i=1}^{N}U_i^2\right)\left(\sum_{i=1}^{N}V_i^2\right) \qquad \ldots(2)$$

if and only if for some $k \in \mathbf{R}$, $V_i=kU_i$ for each $i=1, 2, \ldots\ldots$, N. Taking the positive square roots of each side in (1) and (2), and observing that the ΣU_iV_i may be positive or negative, we get

$$\left|\sum_{i=1}^{N}U_iV_i\right|\leqslant\sqrt{\left[\sum_{i=1}^{N}U_i^2\sum_{i=1}^{N}V_i^2\right]},$$

and $$\left|\sum_{i=1}^{N} U_i V_i\right| = \sqrt{\left(\sum_{i=1}^{N} U_i^2 \sum_{i=1}^{N} V_i^2\right)}$$ if and only if $V_i = kU_i$ for each i and some $k \in \mathbf{R}$.

Division by the *positive* number $\sqrt{\left(\sum_{i=1}^{N} U_i^2 \sum_{i=1}^{N} V_i^2\right)}$

now gives

$$\frac{|\Sigma U_i V_i|}{\sqrt{[\Sigma U_i^2 \Sigma V_i^2]}} \leqslant 1,$$

equality holding if and only if $V_i = kU_i$ for each i and some k. Thus

$$|r| \leqslant 1,$$

equality holding if and only if $V_i = kU_i$ for each i and some k. Thus in general $-1 \leqslant r \leqslant 1$. Also, $|r| = 1$ if and only if $V_i = kU_i$, but that is precisely the condition that the points (U_i, V_i) should lie on the line $Y = kX$.

EXERCISE 13 (*b*)

[*Calculate all answers correct to two decimal places.*]

1. In each of the following pairs of the attributes, if you expect a linear correlation then say '+', or '−' according as you think it to be positive or negative. Indicate no correlation by a '0'.
   (*a*) Rainfall and attendance at a Test match.
   [illegible] The age of a common building and its strength.
   (*c*) Price of [illegible] and price of bun.
   (*d*) Amount of hard work [illegible] performance at examination (upto a reasonable level !).
   (*e*) Amount of unemployment and wage rates.
   (*f*) Cost of life insurance and age of an adult at which it is bought.
   (*g*) Beauty and brain.
2. The following table gives a comparison between weights in kilogrammes of ten boys and their marks in a test. Draw a scatter diagram. Is there an evidence of linear correlation ?

Weight (in kg)	45	47	50	51.5	55	57	60.5	61	62	63
Marks	22	20	26	15	8	25	28	10	12	14

3. The following table gives maximum temperature in degrees centigrade on seven days and the number of bottles of cold drinks sold at a shop on these days. Draw a scatter diagram and state the type of correlation if there exists any.

Temperature	25	23	25	21	16	18	20
No. of bottles sold	42	39	41	27	5	16	25

4. A spiral spring is loaded with weights and extensions caused are noted. The table below gives the recorded data. Draw a scatter diagram and state the type of correlation if there exists any.

Weight (in kg)	0	0.5	0.10	0.15	0.20	0.25	0.30
Extension (in cm)	0	10.0	20.4	30.4	41.0	50.6	61.2

5. A stone is falling from rest. The distances covered are recorded. The following table gives the data. Draw a scatter diagram and comment on the correlation if there exists any.

Time (in seconds)	0	0.5	1	1.5	2	2.5	3	3.5
Distance (in metres)	0	1	5	10	19	30	43	58

6. For each of the problems from 2 to 5 above, guess the value of Pearson's correlation coefficient. Calculate actually and compare your results. How is the thickness of the band of dots related to the degree of r. What kind of correlation do you expect between *these* two ?

7. Given the data $\Sigma X=75$, $\Sigma Y=25$, $\Sigma X^2=92{\cdot}25$, $\Sigma Y^2=582{\cdot}25$ and $\Sigma XY=118{\cdot}75$ for 100 pairs of data points (X, Y), calculate the correlation coefficient.

8. Calculate the Karl Pearson's coefficient of correlation between the marks in Physics and Chemistry obtained by 7 students :

Marks in Physics	25	17	21	23	12	18	22
Marks in Chemistry	22	20	24	25	17	11	18

(*D.B.S.S.C.E., 1987*)

9. Calculate the correlation coefficient between the price and consumption for the following data :

Price	5	5.50	6	6.50	7	7.50	8
Consumption	10	10	8	7	7	6	6

(*D.B.S.S.C.E., 1986 C*)

10. Calculate the Karl Pearson's coefficient of correlation between X and Y for the following data :

X	28	40	41	35	38	33	40	36	32	33
Y	23	33	34	30	34	26	28	36	31	38

(*A.I.S.S.C.E., 1985*), (*D.B.S.S C.E., 1984*)

11. Calculate the coefficient of correlation between X and Y for the following data :

X	1	2	3	4	5	6	7	8	9	10
Y	15	11	13	13	12	12	9	10	10	9

(*A.I.S.S.C.E. 1984*)

12. *A certain coral reef fish is capable of changing its sex.* An experiment was made on 11 groups of this fish. A certain number (X) of male fish were removed and the number (Y) of female fish changing sex was recorded. Plot the data of these experiments given below in a scatter diagram. Does there seem to be a linear correlation ? Calculate Pearson's *r* and compare with your conjecture.

X Males removed	3	4	3	5	5	4	6	6	7	9	8
Y Sex changes	2	2	3	4	5	6	5	7	9	9	9

13. Two judges give the following scores to various competitors in a debating competition :

By Judge A	3	2	5	7	8	7	6	5	4	1
By Judge B	7	4	2	3	8	8	7	3	8	6

What have you to say about the *competence* of the judges where evaluating a debating competition is concerned ? Calculate r and verify your judgement (of the judgement of the judges).

14. Find the coefficient of correlation between the ages (in years) and the sum (in rupees) donated by 170 persons for a charity cause, as given by the table below :

Age group \ Sum donated (Rs)	50	100	200	500	1000
15-24	18	20	6	2	–
25–34	21	26	6	5	1
35-44	10	9	3	6	1
45-54	7	8	5	4	–
55-64	8	3	1	–	–

15. Find the coefficient of correlation between X and Y given by the following table :

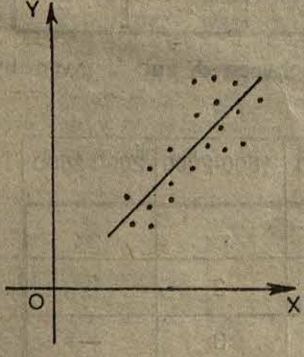
Y \ X	1-500	500-1000	1000-1500	1500-2000	2000-2500
0-200	12	6	—	—	—
200-400	2	18	4	2	1
400-600	—	4	4	3	—
600-800	—	1	—	2	1
800-1000	—	—	—	1	2

13·4. LINEAR REGRESSION

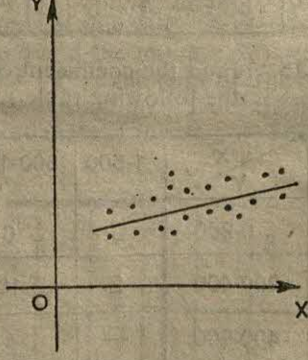
Do you like stories ? You are about to hear one. But first a question. Do you think intelligent parents have intelligent sons ? fat parents, fat sons ? tall parents, tall sons ? Well a certain British scientist, Sir Francis Galton, got to thinking about such questions. He made extensive studies as to how mental and physical traits are passed on to children from parents, from generation to generation. One of his studies concerned the height of sons (he treated daughters as sons by multiplying their heights with 1·08) of tall (short) parents.

From several groups of people, Galton collected data on the mean height (X) of mother and father, and height (Y) of sons. He arranged his data for each group as a continuous bivariate distribution, and for each class corresponding to the mean height of parents, he computed the average height of sons. In other words, he compiled the conditional distribution of Y on X, where Y and X respectively denote now the mid-values of the various classes of Y (height of sons) and X (mean height of parents). Representing the parents' heights on the X-axis and those of sons on the Y-axis, he drew the scatter diagram of this distribution for each group. And there ! every group showed a linear pattern of dots. (What type of value, of Pearson's r would you expect, high or low ?)

If tall parents had sons *as tall* and short parents sons *as short*, X and Y would have been nearly equal. This means that the various dots in Galton's data in this case would have been quite close to the line Y=X. However, the lines close to Galton's dots were a little tilted towards the x-axis. But this meant that in



(*a*)



(*b*)

Fig. 13·5.

general, short parents had sons a bit taller, and the tall parents had sons a bit less tall than themselves. In other words, the heights of

sons showed a tendency to *regress* (revert) toward the mean height. This is where the word *regression* comes from.

Galton used the lines close to the dots to represent his data. Even though the dots did not exactly lie on the lines, they were close enough to be treated as such. The greater the size of the group, the nearer the dots to the line. Thus it was to be expected that if data were collected from *all the parents all over*, the dots would almost lie on these lines. Such lines were called by Galton the *lines of regression*. These lines served another important purpose in addition to the general conclusion drawn above regarding the regression of heights towards the mean height. For a given mean height of parents, one could find from this line, the *approximate* height of their sons. One merely had to read the corresponding Y value from the line for a given value of X. Thus lines of regression were used to *estimate* or *predict* the heights of sons from the heights of the parents.

It must be obvious, that the story does not end with Galton's experiments on the stature of sons. The technique used by Galton can be used easily in numerous other situations ; in fact anywhere, where Pearson's r shows a strong linear relationship. Notice that by itself, *r only gives an indication of linear relationship ; it does not tell you how to* use this relation in order to *determine one variable from the other*. This determination is effected by means of regression techniques like Galton's. For given values of one variable, we can *estimate* or *predict* the corresponding values of the other by using a *line of regression* similar to that of Galton. The only puzzling piece perhaps is the choice of a line from amongst the several which closely fit the dots. Very well ! we shall not play the Auntie cat who kept away from the lion the art of climbing a tree. Let us tell you *how to fit a line to a given scatter diagram so as to be able to use it as a regression line in order to* **estimate** *or* **predict** *the values of one variable from those of the other.*

13·4·1. The Principle of Least Squares

As has been mentioned time and again, our limited data generally concern a sample. On the basis of this sample, we wish to draw conclusions regarding the whole population of which the sample is a part. In case of bivariate data, our interest, often, is in discovering what values of one variable are to be expected for certain given values of the other. For example, we may wish to know how the yield (Y) of an agricultural product, say rice, changes when we apply varying amounts (X) of a certain fertilizer, keeping other conditions fixed. Thus we control the variable X giving it certain fixed values $X_1, X_2, \cdots\cdots, X_n$, and observe the resulting yields $Y_1, Y_2, \cdots\cdots, Y_n$. Notice that $X_1, X_2, \cdots\cdots, X_n$ need not be distinct.

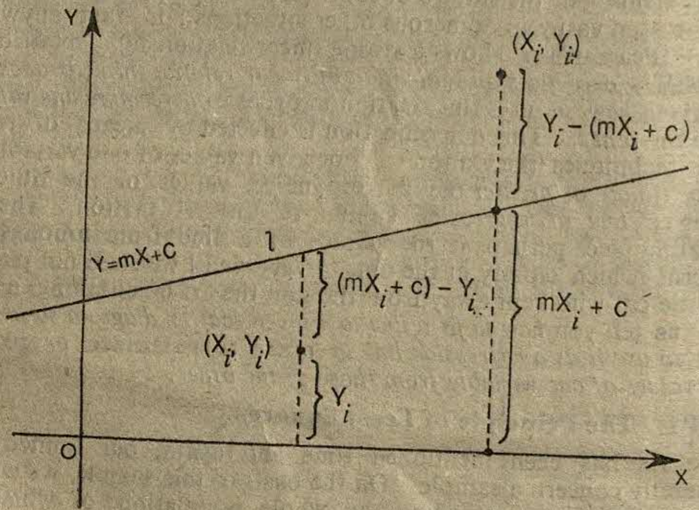
The first step in the process of estimation is to plot the scatter diagram of the data and see if the dots form a linear pattern. In case they do, we would like to find that linear relation which best

approximates the situation. (Remember, unless there is perfect correlation, there does not exist a functional relation as such between X and Y !). In other words, the idea is to fit a line to the data as best as we can, and to predict the values of Y corresponding to given values of X from the equation of this line.

Suppose that the equation of the *line l of best fit*, or the *prediction line*, or the line which we wish to use as the line to predict the values of Y for given values of X, is $Y=mX+c$. Then for $X=X_i$, we shall predict the value of Y as mX_i+c. But the actual observed value is Y_i. Thus

(observed value of Y)—(predicted value of Y)$=Y_i-(mX_i+c)$.

Notice that this is the *error* we shall be committing in predicting Y corresponding to X_i from the line l and it is given by the vertical distance of the dot (X_i, Y_i) from the line l. Clearly, the line l should be such as to minimize the total error. Now denoting the observed value by O_i and the predicted value by P_i, the difference



Fig, 13·6.

O_i-P_i is positive for dots above the line and negative for points below it. Hence if we take the simple sum $\Sigma(O_i-P_i)$ of these deviations, some of the positive errors will cancel out some of the negative errors and we shall have no indication of the amount of total error from the value of the sum. This difficulty can be surmounted by taking the absolute values $|O_i-P_i|$ instead of (O_i-P_i), but then algebraic manipulations become difficult. The relief comes from the *principle of least squares*. According to this principle, that line for which the sum $S=\Sigma(O_i-P_i)^2$ of the deviations

of observed values O_i from the predicted values P_i is a minimum, is taken as the *line of best fit* and *is used to predict the values of Y from those of X.*

13·4·2. Line of Regression of Y on X

In the above discussion, the variable X which is controlled by the experimenter, is known as the **conditioning** or the **causal** (not casual, please !) or the **input** or the **predictor** or the **independent** variable. Y is known as the **response** or the **output** or the **dependent** variable. The line l is known as the **line of regression of Y on X**. The equation of this line involves the two parameters m and c. For different values of m and c, the sum $S(=\Sigma(O_i-P_i)^2)$ would be different. Hence m and c are to be so chosen that S is minimum. There are simple techniques to find m and c for which S has the least value, but unfortunately, the required mathematical tools are as yet unknown to you. For reasons you will learn in higher classes, the values of m and c are obtained by solving the two linear equations (1) and (2) below. These equations are known as the *normal equations.**

$$\Sigma(Y_i-mX_i-c)=0 \qquad \text{...(1)}$$
$$\Sigma X_i(Y_i-mX_i-c)=0 \qquad \text{...(2)}$$

Simple algebraic manipulations on (1) and (2) yield

$$m=S_{XY}/S_{XX},$$
$$c=\bar{Y}-(S_{XY}/S_{XX})\bar{X}.$$

Recall that
$$S_{XY}=\Sigma(X-\bar{X})(Y-\bar{Y}),$$
$$S_{XX}=\Sigma(X-\bar{X})^2,$$
$$S_{YY}=\Sigma(Y-\bar{Y})^2.$$

Hence the required line of best fit is

$$Y=\frac{S_{XY}}{S_{XX}}X+\left(\bar{Y}-\frac{S_{XY}}{S_{XX}}\bar{X}\right),$$

or

$$Y-\bar{Y}=\frac{S_{XY}}{S_{XX}}(X-\bar{X}).$$

Remark. The line of regression of Y on X passes through $(\bar{X}, \bar{Y})$.

13·4·3 Coefficient of Regression of Y on X

The slope S_{XY}/S_{XX} of the line of regression of Y on X is known as the **coefficient of regression of** Y **on** X, and is also denoted by b_{YX}. Being the slope (of the regression line), it measures change in the value of Y corresponding to a unit change in the value of X. To be precise, change X to X+1 in the equation of the

*To obtain (1) differentiate S, *i.e.*, $\Sigma(Y_i-mX_i-c)^2$ with respect to c, treating X_i, Y_i and m as constant. To get (2), differentiate S w.r.t. to m, treating X_i, Y_i and c as constant.

regression line. This gives the new value of Y as $m\ (X+1)+c$ or $mX+c+m$. The original value was $mX+c$. Thus a change of one unit in the value of X, causes a changeof m units in the value of Y. Thus *the regression coefficient of Y on X is an indicator of change in the Y-value due to a unit change in the X-value.*

We can use any of the formulae given below to calculate the regression coefficient b_{YX}.

$$b_{YX}=S_{XY}/S_{XX} \qquad \text{...(1)}$$

$$=\frac{\Sigma(X-\bar{X})\,(Y-\bar{Y})}{\Sigma(X-\bar{X})^2} \qquad \text{...(2)}$$

$$=\frac{\text{COV}(X,\,Y)}{\sigma_X{}^2}, \qquad \text{...(3)}$$

where $\frac{1}{N}\Sigma(X-\bar{X})\,(Y-\bar{Y})$ is the covariance of X and Y. Also, since

$$r=\frac{S_{XY}}{\sqrt{(S_{XX}S_{YY})}},$$

$$b_{YX}=\frac{r\sqrt{(S_{XX}\,S_{YY})}}{S_{XX}}$$

$$=r\sqrt{\left(\frac{S_{YY}}{S_{XX}}\right)} \qquad \text{...(4)}$$

$$=\frac{r\sigma_Y}{\sigma_X} \qquad \text{...(5)}$$

For the purpose of calculation, the following expressions for b_{YX} may be the most convenient ;

$$b_{YX}=\frac{\Sigma XY-N\bar{X}\bar{Y}}{\Sigma X^2-N\bar{X}^2} \qquad \text{...(6)}$$

$$=\frac{\Sigma X'Y'-\frac{1}{N}\,\Sigma X'\Sigma Y'}{\Sigma X'^2-\frac{1}{N}\,(\Sigma X')^2}, \qquad \text{...(7)}$$

where $X'=(X-a)/h$, $Y'=(Y-b)/h$ for suitable real numbers a, b and h.

Remarks. 1. In view of (5), the line of regression of Y on X can be written as

$$Y-\bar{Y}=\frac{r\sigma_Y}{\sigma_X}\,(X-\bar{X}).$$

2. b_{YX} has the same sign as r. (Why ?)

Example 8. *The following table gives the dosage (X) of a certain medicine (in mg) and the number (Y) of hours the patient is relieved of pain. Find the line of regression of Y on X. What is the expected duration of relief for a dosage of (a) 5 mg, (b) 6·5 mg? Also find the regression coefficient of Y on X.*

X (*dosage*)	3	3	4	4	5	6	7	8
Y (*duration of relief*)	9	8	10	9	14	16	18	20

Solution. Here the predictor variable is X, the dosage. Y, the number of hours of relief, is the variable to be predicted. The dot-pattern in the scatter diagram shows linear correlation. Hence we fit a least square line to the data. The necessary computations are shown in the table below :

X	$X-\bar{X}$	$(X-\bar{X})^2$	Y	$Y-\bar{Y}$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
3	−2	4	9	−4	16	8
3	−2	4	8	−5	25	10
4	−1	1	10	−3	9	3
4	−1	1	9	−4	16	4
5	0	0	14	1	1	0
6	1	1	16	3	9	3
7	2	4	18	5	25	10
8	3	9	20	7	49	21
Total 40		24	104		140	59
$\bar{X}=5$		S_{XX}	$\bar{Y}=13$		S_{YY}	S_{XY}

Hence the line of regression of Y on X is

$$Y-\bar{Y}=\frac{S_{XY}}{S_{XX}}(X-\bar{X})$$

or
$$Y-13=\frac{59}{24}(X-5)$$

or
$$Y=2{\cdot}46\times-12{\cdot}30+13$$
$$=2{\cdot}46++0{\cdot}70. \qquad \ldots(1)$$

(*a*) When the dosage X is 5 mg, the *predicted* value of Y is $(2{\cdot}46 \times 5 + 0{\cdot}7)$ *h*, or 13 hours.

(*b*) When the dosage is 6·5 mg, the *estimated* number of hours of relief is $2{\cdot}46 \times 6{\cdot}5 + 0{\cdot}7$, or 16·69.

Also, the required regression coefficient is S_{XY}/S_{XX}, 2·46 approximately.

Remarks. 1. Notice that when the dosage is 5 mg, the estimated number of hours of relief is 13, whereas the observed number of hours of relief is 14. This should not upset you. In fact it is a reminder to you that the *relationships we are talking about here are not functional relations or formulas, they are only ESTIMATES in the sense that if we repeated our experiments a large number of times, or had a very large sample, then the observed values would be nearer to the estimated value 13 than to the observed value 14.* As a matter of fact, 14 is a value obtained as an outcome of one experiment or a sample of size 1. This small sample cannot give us a correct picture of the whole population.

2. Another interesting point that this example shows us is that regression lines can be used to predict values for those values of the input variable also which were not a part of our experiment or sample but *which fall within the range of values given to the input variable.* Thus even though we experimented with some values of X ranging from 3 to 8, 6·5 was not such a value. Yet we are able to have a valid guess about the number of relief-hours such a dosage should provide. However, a word of caution is needed here. The value 6·5 was very much a value within the range of values included in our experiment and we could use the regression line successfully

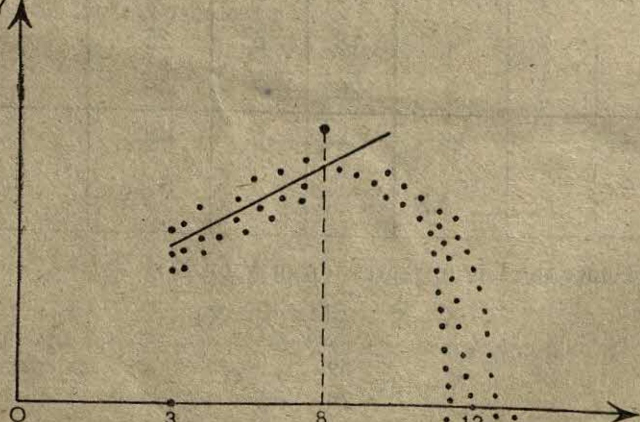


Fig. 13·7.

to obtain in estimated value of Y for $X=6\cdot5$. Had we taken a value like 20, the predicted value might have been far from the actual values. For example, Suppose the dot pattern had been like the one shown Fig. 13·7. It appears that the duration of relief increases with an increase in dosage in the range 3 mg to 8 mg, and 8 is nearly the optimum dosage. A further increase results in over-medication perhaps and duration of relief steadily decreases in the range 8 mg to 12 mg, after which the medicine really has an adverse effect. Thus a dosage like 20 mg would be very injurious to the patient whereas were we to use the regression line fitted to the data in the range 3 mg to 8 mg, we shall predict a very long duration (50 hours nearly) of relief. The moral is that regression lines may be used to *interpolate* (predict within the given range) but NEVER to *extrapolate* (predict outside the given range).

3. There is another pit-fall in addition to extrapolating. You might feel like using the regression line obtained above to predict the dosage required, if you wish to provide relief for a certain duration, say whole night, or nine hours. Now nine hours is well within the range of durations of our experiment. Yet it would not be judicious to use this line for predicting the dosage from the durations of relief. Do not forget that this line was obtained by minimizing the errors (the vertical distances) in the predicted and the observed values of Y. If we wish to predict X, then we shall have to minimize the errors in observed values of X and the estimated values of X. There errors are given by the horioznta1 distances of the dots from the line of best fit. When the coefficient of correlation is not very high so that the dots are scattered in a not-very-thin-band, the lines fitted to the data by minimizing the sums of squares of horizontal and vertical distances may be quite different as Example 10 shows.

13·4·4. Line of Regression of X on Y

The line fitted to the data by minimizing the sum of squares of X-errors (*i.e.*, the horizontal distances) instead of Y-errors is known as the **line of regression of X on Y**. This line is used to predict the values of X from those of Y. The equation of this line turns out to be

$$X-\bar{X}=\frac{S_{XY}}{S_{YY}}(Y-\bar{Y}),$$

or

$$X-\bar{X}=\frac{r\,\sigma_X}{\sigma_Y}(Y-\bar{Y})$$

The coefficient of regression of X on Y, denoted by b_{XY}, is $r\sigma_X/\sigma_Y$ or $r\sqrt{(S_{XX}/S_{YY})}$, and hence has the same sign as r. Here, Y is the input or the conditioning variable and X the dependent variable.

Notice that the product of the two regression coefficients turns out to be r^2. Hence *the correlation coefficient is the geometric*

mean of the regression coefficients, and has the same sign as either regression coefficient.

Example 9. *The following table gives some figures regarding the batch-size (X) of production (in thousands) and the cost of a ball (in rupees).*

X (batch-size in thousands)	2	3	4	5	6	7	8	9	10	11
Y (cost per ball in rupees)	10	9	8	8	8	8	6	5	6	7

Obtain the correlation coefficient and the two lines of regression. What would be the expected cost of a ball for batch-size 5500? What batch-size might lead to a cost of Rs. 9·50 per ball?

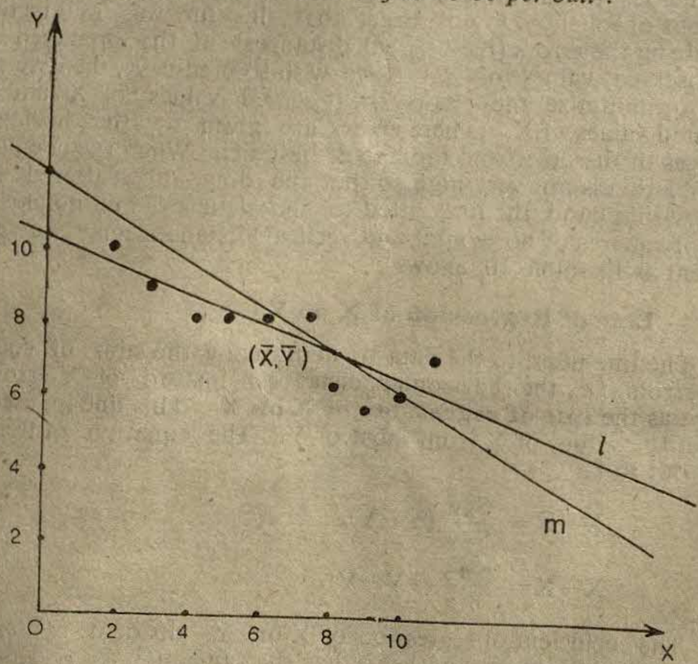


Fig. 13·8.

Solution. Here $\bar{X}=6{\cdot}5$, $\bar{Y}=7{\cdot}5$, $\sigma_X=2{\cdot}87$, $\sigma_Y=1{\cdot}43$, and $r=-{\cdot}84$. The line of regression of Y on X is

$$Y-\bar{Y}=\frac{r\,\sigma_Y}{\sigma_X}(X-\bar{X}),$$

or $$Y-7{\cdot}5=-{\cdot}42\,(X-6{\cdot}5),$$

or $$Y=-0{\cdot}42X+10{\cdot}23. \qquad ...(1)$$

This is the line l in Fig. 13·8.

This line of regression of X on Y is

$$X-\bar{X}=\frac{r\sigma_X}{\sigma_Y}(Y-\bar{Y}),$$

or $$X-6{\cdot}5=-1{\cdot}69\,(Y-7{\cdot}5),$$

or $$X=-1{\cdot}69\,Y+19{\cdot}18. \qquad ...(2)$$

This is the line m in Fig. 13·8.

To obtain the estimated cost per ball (Y) for a given batch-size (X) of 5500, we use the line of regression of Y on X. Since X is in thousands, substituting $X=5{\cdot}5$ in (1), we get

$$X=-0{\cdot}42\times5{\cdot}5+10{\cdot}23=7{\cdot}92$$

Thus the estimated cost per ball for a production run of 5500 balls is Rs. 7·92.

To estimate the batch-size for a given cost, we must use the prediction line of X on Y. Thus if the cost (Y) is Rs. 9·50, the estimated batch-size (X) is given by (2). Substituting $Y=9{\cdot}5$ in (2), we get

$$X=-1{\cdot}69\times9{\cdot}5+19{\cdot}18=3{\cdot}12.$$

Thus to the nearest thousand, a batch-size of 3000 would produce a cost Rs. 9·50 per ball.

Remarks. 1. The estimated values can be read from the graph of the line also, if the graphing is accurately done.

2. To plot a line, only two points on the line are required. Since each of the regression lines passes through $(\bar{X},\bar{Y})$, only one more point is required. This point might be taken as the point where the line meets a co-ordinate axis.

3. In practice, we talk of only about the line of regression of Y on X since either of the two variables may be labelled as X. We label the input variable or the conditioning variable as X and mark it on the X-axis. The dependent variable, or the variable to be estimated is labelled Y and shown on the Y-axis.

Example 10. *Draw the scatter diagram and plot the two regression lines for the following bivariate data* :

X	3	4	4	5	6	7	8
Y	3	6	10	4	8	9	6

Solution. Here the various statistics of the two variables are

$$\bar{X}=5.29,\ \bar{Y}=6.57,\ \sigma_X=1.67,\ \sigma_Y=2.38,\ r=0.31.$$

Thus line l of regression of Y on X is

$$Y-\bar{Y}=\frac{r\,\sigma_Y}{\sigma_X}(X-\bar{X})$$

or $$Y-6.57=0.44(X-5.29),$$

or $$Y=0.44\,X+4.23. \qquad \ldots(1)$$

The line of regression of X on Y is

$$X-\bar{X}=\frac{r\sigma_X}{\sigma_Y}(Y-\bar{Y}),$$

or $$X-5.29=0.22\,(Y-6.57),$$

or $$X=0.22\,Y+3.86. \qquad \ldots(2)$$

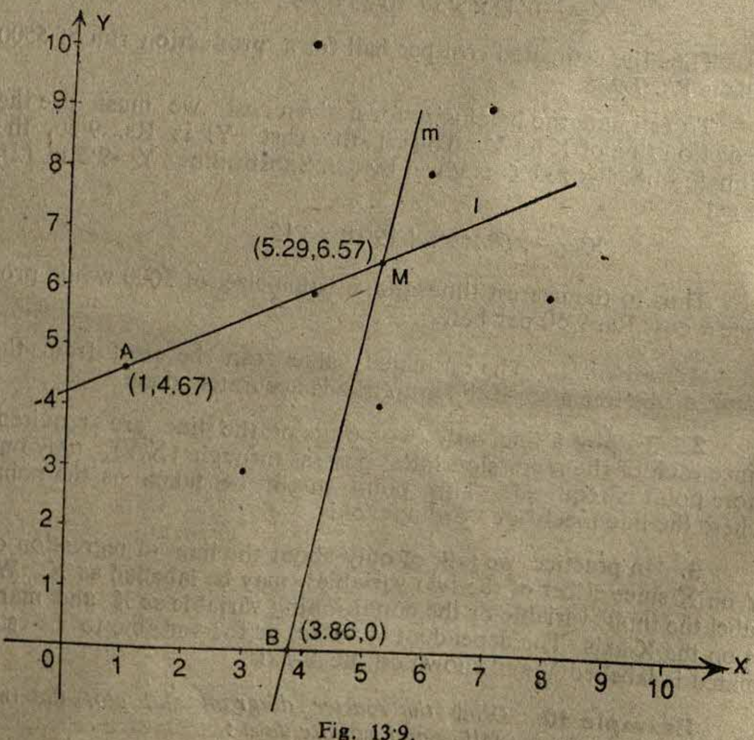


Fig. 13·9.

The scatter diagram shows that the correlation is rather weak. This is also corroborated by the low value of r. To plot the regression lines, we note that both pass through the point $(\bar{X}, \bar{Y})$ which is (5·29, 6·57) in this case. A point on l is A(1, 4·67) and a point on m is B(3·86, 0). Thus AM and BM are the lines of regression of Y an X and of X on Y respectively.

Remarks. 1. For small values of r, the angle between the regression lines is large. The higher the value of r, the smaller this angle and closer the lines. In case $r=1$ (perfect correlation), the two lines of regression coincide with the line whose equation is the functional relation connecting X and Y.

2. For a particular value of Y, say 7, the estimated value of X obtained from (2), the line of regression of X on Y is 5·4. This is quite different from the value 6·3 of X obtained from (1) for Y=7. Thus it is important to use the correct line for prediction.

13·4·5. Error of Prediction

The line of regression of Y on X (*i.e.*, the line to predict Y from values of X) was so fitted that the sum S of the squares of the deviations of the observed value of Y from the estimated values of Y were as small as possible. The predicted values are only rather good estimates and differ in general from the actual observed values. This deviation from the observed values is a kind of *error* arising due to prediction. To have an idea of that error, we use a measure based on all these deviations It is known as the **standard error of prediction** (SEP) and is defined as

$$E_{YX}=\left[\frac{1}{N}\sum (Y-Y_p)^2\right]^{\frac{1}{2}}, \qquad ...(1)$$

where Y_p stands for the predicted value and Y the actual value. Since the equation of the line of regression of Y on X is

$$Y-\bar{Y}=\frac{r\sigma_Y}{\sigma_X}(X-\bar{X}),$$

therefore, the predicted value Y_p of Y is given by

$$Y_p=\bar{Y}+\frac{r\sigma_Y}{\sigma_X}(X-\bar{X}),$$

Hence $$E_{YX}=\left[\frac{1}{N}\sum (Y-Y_p)^2\right]^{\frac{1}{2}},$$

$$=\left[\frac{1}{N}\sum\left\{Y-\bar{Y}-\frac{r\sigma_Y}{\sigma_X}(X-\bar{X})\right\}^2\right]^{\frac{1}{2}},$$

$$=\left[\frac{1}{N}\sum\left\{(Y-\bar{Y})^2-\frac{2r\sigma_Y}{\sigma_X}(X-\bar{X})(Y-\bar{Y})+\frac{r^2\sigma^2_Y}{\sigma_X^2}(X-\bar{X})^2\right\}\right]^{\frac{1}{2}},$$

$$=\left[\frac{1}{N}\left\{\sum(Y-\bar{Y})^2-\frac{2r\sigma_Y}{\sigma_x}\sum((X-\bar{X})(Y-\bar{Y}))\right.\right.$$

$$\left.\left.+\frac{r^2\sigma_Y^2}{\sigma_X}\sum(X-\bar{X})^2\right\}\right]^{\frac{1}{2}}$$

$$=\left[\sigma_Y^2-\frac{2r\sigma_Y}{\sigma_X}.\,\sigma_X\,\sigma_Y r+r^2\,\sigma_Y^2\right]$$

$$=\sigma_Y\sqrt{(1-r^2)}. \qquad \cdots(2)$$

We may similarly calculate E_{XY}, the SEP of X on Y and show that it is equal to $\sigma_X\sqrt{(1-r^2)}$. Thus

$$E_{XY}=\sigma_X\sqrt{(1-r^2)}.$$

Remarks. 1. SEP is quite similar to standard deviation (SD). The only difference is that whereas in case of SD, the deviations are taken from the mean, in case of SEP, they are taken from the regression line.

2. $SEP \geqslant 0$.

13·4·6. Relation between r and the Standard Error of Prediction

Can you say what is the maximum possible predictive error and what is the minimum ? Since $E_{YX}=\sigma_Y\sqrt{(1-r^2)}$, the minimum error is zero, and corresponds to $r=\pm 1$. That, of course, is as it should be. In case of perfect correlation, the data points all lie on the lines of regression, the predicted values are the same as the actual values and there is no error. The maximum value of SEP of Y on X is obtained for $r=0$ and thus it is σ_Y. Similarly, the other maximum SEP is σ_X. This is again not surprising. When $r=0$, the lines of regression reduce to $Y-\bar{Y}=0$ and $X-\bar{X}=0$. Thus no matter what is X, the corresponding predicted value of Y is $\bar{Y}$. Hence $E_{YX}^2=\frac{1}{N}\Sigma\,(Y-Y_p)^2=\frac{1}{N}\Sigma\,(Y-\bar{Y})^2=\sigma_Y^2$. Similarly, E_{XY} $=\sigma_Y$ in this case. (Of course, it would be stupid to use the lines of regression for prediction in case of no correlation !) Having seen the extreme cases, let us examine what happens in between.

For $r=0$, $E_{YX}=\sigma_Y$ and for $r=1$, $E_{YX}=0$. For $r=0{\cdot}5$, would E_{YX} be $1/2\ \sigma_Y$? No, of course not ! E_{YX} *and r do not vary in direct proportion*. For $r=0{\cdot}5$,

$$E_{YX}=\sigma_Y\sqrt{(1-0{\cdot}25)}=0{\cdot}87\sigma_Y.$$

Thus as r reduces by 50%, E_{YX} reduces only by 13% ; 87% of the predictive error is still present. The table on page 741 gives some values of r together with the proportionate SEP and the corresponding decrease in it. The table shows that as r increases from 0 to 1 SEP decreases slowly until r becomes greater than 0·6.

r	1.00	0.90	0.80	0.75	0.60	0.50	0.40	0.25	0.00
SEP	0%	44%	60%	66%	80%	87%	92%	97%	100%
Reduction in SEP	100%	56%	40%	34%	20%	13%	8%	3%	0%

We can use this table with advantage to decide for how high values of r, should we bother to predict at all. For example, even a moderate value like 0·5 of r has a high margin of error, so that we would not be wise to place much faith in our predictions for such a value of r.

Example 11. *The standard error of prediction of Y on X for the data of the previous example is*

$$\sigma_Y\sqrt{(1-r^2)}=2{\cdot}38\sqrt{(1-0{\cdot}31)^2}=2{\cdot}38\times0{\cdot}96=2{\cdot}28.$$

Also, SEP of X on Y

$$\sigma_X\sqrt{(1-r^2)}=1{\cdot}67\times0{\cdot}96=1{\cdot}6$$

We can verify these values by actually calculating the quantities $\left[\frac{1}{N}\Sigma(Y-Y_p)^2\right]^{1/2}$ and $\left[\frac{1}{N}\Sigma(X-X_p)^2\right]^{1/2}$ etc. The equation of the line of regression is

$$Y=0{\cdot}44X+4{\cdot}23. \qquad \cdots(1)$$

Denoting by Y_a the predicted value of Y obtained from (1) for $X=a$, we get

$$Y_3=5{\cdot}55,\ Y_4=5{\cdot}99,\ Y_5=6{\cdot}43,\ Y_6=6{\cdot}87,\ Y_7=7{\cdot}31,$$

and $\quad Y_8=7{\cdot}75.$

Note that since X is taking consecutive values and b_{YX} measures change in Y corresponding to a unit change in X, the successive values of Y_a can be obtained by merely adding the number 0·44 (the regression coefficient b_{YX}) to the previous Y_a value.

When X=3, the actual value of Y is 3 and the predicted value of Y is $Y_3(=5{\cdot}55)$, Similarly, for X=4, the actual y-value is 6 and the predicted Y value is $Y_4(=5{\cdot}99)$, and so on. Hence

$$\begin{aligned}\Sigma\,(Y-Y_p)^2&=(3-5{\cdot}55)^2+(6-5{\cdot}99)^2+(10-5{\cdot}99)^2\\&+(4-6{\cdot}43)^2+(8-6{\cdot}87)^2+(9-7{\cdot}31)^2\\&+(6-7{\cdot}75)^2,\\&=35{\cdot}68.\end{aligned}$$

Hence $\quad E_{YX}=\left[\frac{1}{N}\,\Sigma\,(Y-Y_p)^2\right]^{1/2}=2{\cdot}26.$

The earlier value obtained from the formula $E_{YX}=\sigma_Y\sqrt{(1-r^2)}$ is 2·28. The slight difference in the value is due to the fact that we are taking approximations. You can similarly verify the correctness of the other SEP.

EXERCISE 13 (*c*)

1. For the data given below :

X	2	2	3	3	4	4
Y	8	7	7	6	5	6

(*a*) Plot the scatter diagram.

(*b*) Calculate $\bar{X}$, $\bar{Y}$, S_{XX}, S_{YY}, σ_X, σ_Y, and r.

(*c*) Calculate the regression coefficients b_{XY} and b_{YX}.

(*d*) Find the line of regression of Y on X and draw it on the scatter diagram. Verify that $(\bar{X}, \bar{Y})$ lies on it.

(*e*) Predict the value of Y for $X=3·5$.

(*f*) Calculate E_{YX}, the standard error of prediction of Y on X by using the formula $\sigma_Y(1-r^2)^{1/2}$. Verify the correctness of your result by actually evaluating the expression

$$\left[\frac{1}{N}\Sigma\,(Y-Y_p)^2\right]^{1/2}.$$

(*g*) Verify that $\Sigma\,(Y-Y_p)=0$.

2. The following summary statistics are available for some bivariate data for which the assumption of linear regression is valid ;

$$N=10,\ \bar{X}=3·6,\ \bar{Y}=2·4,\ S_{XX}=10·8,\ S_{XY}=2·8$$
and $\quad S_{YY}=1·5$.

(*a*) Calculate the regression coefficients.

(*b*) Obtain the correlation coefficient as the geometric mean of the regression co-efficients. Verify by some other formula.

(*c*) Find the line of regression of X on Y. Verify that $(\bar{X}, \bar{Y})$ lies on it. Which is the input variable, X or Y ?

(*d*) Find E_{XY}, the standard error of prediction of X on Y.

3. In twenty similar experiments, data on the strength (Y) of plastic fibre and the size (X) of the drops of a mixing chemical in suspension are collected, and the following statistics result :

$$\bar{X}=8,\ \bar{Y}=54,\ \sigma_X=0·53,\ \sigma_Y=1·39,\text{ and } S_{XY}=-12.$$

(*a*) Can linear correlation be assumed ? Is correlation positive ?

(*b*) Obtain the line of regression of Y on X.

(c) Obtain the expected fibre strengths for drops of size 9 and 10.

4. The following summary statistics were obtained from the heights-at-age-ten (X) of 1000 boys and their heights-at-age-twenty (Y):

$\bar{X}=116$ cm, $\bar{Y}=162$ cm, $\sigma_X=7.5$, $\sigma_Y=10$, $r=0.7$

(a) Find a suitable regression equation to predict heights-at-age-twenty from those at age 10.

(b) Use the above equation to predict the heights at age 20 of the following ten-year-old boys:

Name	Anand	Prakash	Chetan
Height (in cm)	118	114	116

[Did you *calculate* Chetan's height ?]

(c) Can we use the above line to predict the heights at age 50 for some boys whose heights at age 10 are known ?

[Do not forget that we cannot predict outside the range of observation but don't divorce common sense either !]

5. The following summary statistics were obtained by observing 37 diabetic patients on blood non-protein nitrogen in mg per cent (Y) and serum chloride in milliequivalents per *l* (X):

$\bar{X}=95.46$, $\bar{Y}=47.19$, $\Sigma X^2=3{,}38{,}984$, $\Sigma Y^2=95{,}248$, and $\Sigma XY=1{,}64{,}079$.

Show that the correlation coefficient is -0.54 and the standard error of prediction of Y on X is nearly 10.

6. The lines of regression of a bivariate data are $X-2Y+1=0$ and $2X-9Y+6=0$. Find the arithmetic means of X and Y. Find Pearson's correlation co-efficient. Which of the two lines is the line of regresssion of Y on X ?

SUMMARY

1. **Bivariate data** are data which involve the observation of two characteristics for every unit of observation.
2. **Bivariate frequency distributions** arise from observations of two variables on the same unit. Theoretically, these are functions which assign to a pair of values, its frequency.
3. A **two-way frequency table** is a tabular representation of bivariate data.
4. The two univariate frequency distributions associated with a bivariate frequency distribution are known as the **marginal distributions** of these data and are obtained by associating the relevant marginal totals with various values of the two variables.

5. The two **conditional distributions** associated with a bivariate frequency distribution are the distributions which assign to various values of either variable the mean values of the other (corresponding to this value).
6. **Correlation** is the term used for the association between the two variables of bivariate data. Correlation does not necessarily mean a cause-effect relation.
7. **Pearson's Correlation Coefficient r** measures the degree of linear correlation between two variables X and Y. It is given by

$$r=\frac{S_{XY}}{\sqrt{(S_{XX}S_{YY})}},$$

where $S_{XY}=\Sigma(X-\bar{X})(Y-\bar{Y})$, $S_{XX}=\Sigma(X-\bar{X})^2$, and $S_{YY}=\Sigma(Y-\bar{Y})^2$. For other formulae, refer to page 715. r is a dimensionless number such that $-1\leq r\leq 1$. Higher (resp lower) values of $|r|$ imply strong (resp. weak) linear correlation. $r=\pm 1$ imply perfect correlation or a linear functional relation between X and Y. Positive values of r indicate **positive correlation**, *i.e.*, the two variables increase (decrease) together in this case. Negative values of r indicate **negative correlation** and in this case one variable increases as the other decreases.
8. **Scatter diagram** is a graphic device to represent quantitative bivariate data with each data point being marked by a *dot* in the two dimensional plane.
9. **Regression.** The word meaning of *regress* is to 'revert', but regression techniques are used to *predict* or *estimate* the values of one quantitative variable from those of another, related somehow with the former.
10. **Principle of least squares** is a rule according to which parameters are so chosen as to make the sum of squares of the deviations of observed values from theoretical values as small as possible.
11. **Regression lines** are the lines of *best fit* to given bivariate data determined according to the principle of least squares.
12. **Line of regression of Y on X is**

$$Y-\bar{Y}=\frac{r\sigma_Y}{\sigma_X}(X-\bar{X})\cdot$$

It is used to predict values of Y from those of X. Here X is the *predictor* or *input* variable and Y the dependent variable.
13. **Line of regression of X on Y** is used to predict values of X from those of Y and has equation

$$X-\bar{X}=\frac{r\sigma_X}{\sigma_Y}(Y-\bar{Y}).$$

14. The quantities $\frac{r\sigma_Y}{\sigma_X}$ and $\frac{r\sigma_X}{\sigma_Y}$ occurring in the lines of regression are known as the **regression coefficients** of Y on X and of X on Y respectively.
15. $E_{YX}=\sigma_Y\sqrt{(1-r^2)}$ is known as the **standard error of prediction of Y on X**. $E_{XY}=\sigma_X\sqrt{(1-r^2)}$ is similarly the **error of prediction of X on R.**

TEST YOUR UNDERSTANDING XIII

1. If Pearson's $r=1$ for some data, then all the dots of the scatter diagram of these data lie on the line
(*a*) $Y=X$, (*b*) $Y=-X$,
(*c*) $Y=mX$ for some m, (*d*) $Y=mX+c$ for some m and c.
2. $r=0{\cdot}2$ indicates that the linear correlation in necessarily
(*a*) positive, (*b*) negative,
(*c*) quite strong, (*d*) rather weak.
3. $r=-{\cdot}97$ indicates that

(*a*) correlation is negative and curvilinear,
(*b*) correlation is linear and negative,
(*c*) concentration of dots is in the third and the fourth quadrant,
(*d*) none of these.

4. Which of the following could have a correlation coefficient equal to $-{\cdot}9$ if its value for the other three is 0, $-{\cdot}8$ and $8{\cdot}5$?

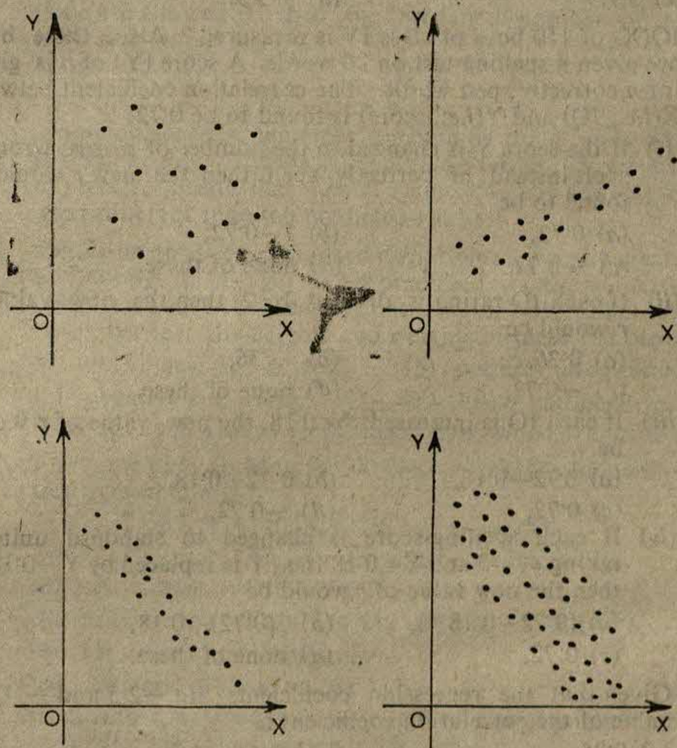


Fig. 13.10.

5. If the correlation coefficient between X and Y is r, then that between Y and X is

(*a*) $-r$ (*b*) $\frac{1}{r}$ (*c*) r, (*d*) $1-r$.

6. Given below are the calculations on 10 pairs of data values (X, Y) :

$$\Sigma(X-\bar{X})^2=25,\ \Sigma(Y-\bar{Y})^2=256,\ \Sigma(X-\bar{X})(Y-\bar{Y})=80,$$

Pearson's r for these data is

(*a*) 0·0125, (*b*) 0·00125, (*c*) 1·0, (*d*) 0·1.

7. Heights and weights were measured in centimeters and grams respectively and Pearson's r between heights and weights was found to be $-\cdot 39$. For the same data, the measurements were changed into meters and kilograms by dividing the figures for heights by 100 and those for weights by 1000 respectively. r was calculated again. The new value of r is

(*a*) $-0\cdot 39$, (*b*) $-\cdot 039$,

(*c*) $3\cdot 9$, (*d*) $-3\cdot 9$.

8. IQ(X) of 150 boys of class IV is measured. Also, these boys are given a spelling test on 20 words. A score (Y) of n is given for n correctly spelt words. The correlation coefficient between X(*i.e.*, IQ) and Y(*i.e.*, score) is found to be $0\cdot 72$.

(*i*) If the score Y is changed to the number of words wrongly spelt instead of correctly spelt, then the new r would be found to be

(*a*) $0\cdot 72$, (*b*) $1-0\cdot 72$,

(*c*) $-0\cdot 72$, (*d*) none of these.

(*ii*) If each IQ rating is divided by 2, then the new value of r would be

(*a*) $0\cdot 36$, (*b*) $-\cdot 36$,

(*c*) $-0\cdot 72$, (*d*) none of these.

(*iii*) If each IQ is increased by $0\cdot 18$, the new value of r would be

(*a*) $0\cdot 72+0\cdot 18$, (*b*) $0\cdot 72-0\cdot 18$,

(*c*) $0\cdot 72$, (*d*) $-0\cdot 72$.

(*iv*) If each spelling-score is changed to standard units by taking $\sigma_Y=3$ and $\bar{Y}=0\cdot 18$ (*i.e.*, Y is replaced by $Y-0\cdot 18/3$), then the new value of r would be

(*a*) $(0\cdot 72-0\cdot 18)/3$, (*b*) $3(0\cdot 72)+0\cdot 18$,

(*c*) $0\cdot 72$, (*d*) none of these.

9. Given that the regression coefficients are $-2\cdot 5$ and $-\cdot 1$, the value of the correlation coefficient is

(*a*) $0\cdot 25$, (*b*) $0\cdot 5$

(*c*) $-0\cdot 5$, (*d*) none of these.

10. Given the values of r, E_{XY} and b_{XY}, we can find the value of

(*a*) $\bar{X}$, $\bar{Y}$, σ_X, (*b*) $\bar{Y}$, σ_X, σ_Y,

(*c*) σ_X, σ_Y, b_{YX}, (*d*) σ_Y, b_{YX}, $\bar{X}$.

11. For some bivariate data, $\Sigma(X-\bar{X})(Y-\bar{Y})=-379\cdot 5$. A possible value of r for these data is

(*a*) $1\cdot 0$, (*b*) $0\cdot 5$,

(*c*) $-\cdot 5$, (*d*) $0\cdot 0009$.

12. The values of r, b_{XY}, b_{YX} and E_{XY} are -10, $-\cdot 9$, $0\cdot 209$ and $-0\cdot 81$ in some order; $0\cdot 209$ must be the value of

(*a*) r, (*b*) b_{XY},
(*c*) b_{XY}, (*d*) E_{XY}.

13. The following statistics were available for some data on the monthly income (X in rupees) of the husbands of 5000 women and the number (Y) of letters in their first name :

$$\bar{X}=3716,\ \bar{Y}=5,\ \sigma_X=1063,\ \sigma_Y=6, \text{ and } r=0.$$

If Susheela's husband earns Rs. 3000 per month (no calculations are allowed !), then the predicted number of letters in her name would be

(*a*) 8, (*b*) 6,
(*c*) 5·5, (*d*) 5.

[Common sense is the most uncommon thing on earth. Even though we know 'Susheela' contains 8 letters, yet we may want to *predict*. Recall that actual values are by and large somewhat different than the predicted values.]

14. Sonu's height is one standard deviation above the mean height of 100 boys of his age. If his predicted weight is one standard deviation above the mean weight of these boys, then the correlation between the heights and weights of these 100 boys is

(*a*) curvilinear and perfect, (*b*) perfect and negative,
(*c*) negative and linear, (*d*) linear and positive.

[It is not such a crazy sum, if only you would bother to find r.]

15. For some data on Monu's 100 class-mates, the following statistics were obtained :

$$\bar{X}=67,\ \bar{Y}=140,\ \sigma_X=7{\cdot}6 \text{ and } \sigma_Y=10{\cdot}3.$$

Monu's X-value is 67. His predicted Y-value is 140.

(*a*) only when $r=1$ or -1 but not for $r=0{\cdot}5$ and $0{\cdot}6$,
(*b*) only when $r=1$ and $0{\cdot}5$ but not for -1 and $-{\cdot}6$,
(*c*) for all of these values of r,
(*d*) for none of these values of r.

[Remember, we always know a certain special point on the lines to predict the values of either variable !]

REVIEW EXERCISE XIII

1. Calculate the coefficient of correlation for the following data between X and Y : (*A.I.S.S.C.E., 1988*)

X	20	25	30	35	40	45
Y	16	10	8	20	5	10

2. Calculate the coefficient of correlation between X and Y for the following data : (*A.I.S.S.C.E., 1987*)

X	1	2	3	4	5	6	7	8	9
Y	2	5	7	8	10	11	10	10	9

3. Calculate the coefficient of correlation between X and Y for the following data : *(D.B.S.S C.E., 1988)*

X	10	7	12	12	9	16	12	18	8	12	14	16
Y	6	4	7	8	10	7	10	15	5	6	11	13

4. Plot the scatter diagram of the following data and in case you think there is linear correlation, calculate the coefficient of correlation.

X	2	3	4	5	6	2	5	3
Y	4	7	7	10	11	6	9	8

5. Two chemical tests were performed on 10 specimens of rocks collected from the moon in order to analyze their composition. Amounts of hydrogen and carbon in parts per million (p.p.m.) as recorded are given below. Calculate Pearson's *r*.

Hydrogen (p.p.m.)	82	90	8	38	20	28	66	20	20	85
Carbon (p.p.m.)	110	99	22	50	50	73	74	77	45	51

6. A student obtained the following values while calculating the correlation coefficient of a bivariate data :

$$\Sigma X = \Sigma Y = 0 \; ; \; \Sigma XY = 185 \; ; \; \Sigma X^2 = 36 \; ; \; \Sigma Y^2 = 968.$$

However, by mistake, he had taken down a data pair (5, 7) as (7, 5). Find the correct correlation coefficient.

7. 20 students are to be subjected to a difficult test. Just before the test, their anxiety is measured by a short-duration test which is known to be quite reliable (*i.e.*, has a high probability of giving correct results). The anxiety measure (X) and scores (Y) in the actual test are given below :

(*a*) Do you think there is *some* relationship (of any kind) between X and Y ?

(*b*) What do you think would be the value of *r*, greater than 0·5 or smaller, when all the 20 students are considered ?

(X) anxiety measure	92	90	89	86	85	82	82	81	80	79
(Y) test-score	38	42	39	45	42	47	49	45	48	51

(X) anxiety measure	77	77	75	74	72	71	70	68	65	63
(Y) test-score	53	47	46	49	46	45	43	41	39	42

(*c*) What do you think would be a rough estimate of the value of *r* for the ten students given in the first table ? Would it be positive ?

(*d*) What about the ten students in the second table ? Do you expect a positive *r* ?

(*e*) Calculate *r* for (*i*) all the 20 students taken together, (*ii*) first ten students, and (*iii*) last ten students. Compare with your conjectures.

[By the way what is the lesson taught by this problem ?]

8. The following statistics were collected during a chemistry experiment to study a possible relationship between output (X) in milligrams and temperature setting (Y) in degree celsius :

$$N=4\ \bar{X}=639,\ \bar{Y}=188,\ \sigma_x^2=210,$$
$$\sigma_Y^2=407,\ \Sigma XY=234.$$

Calculate

(*a*) The correlation coefficient *r*.

(*b*) The line of regression of Y on X.

(*c*) Predict the output when the temperature is set at 60 degrees celsius.

9. To study the effect of the hardness of water on cooking, data were collected on time (Y) taken (in minutes) in boiling rice till soft and the amount (X) of magnesium (milligrams per litre). These are given below :

X	8.5	9	11	8.5	9	11	11	12
Y	21	27	36	23	25	33	35	40

Assuming linear correlation, find the line of regression of Y on X. How much time approximately would it take in boiling

rice if the amount of magnesium in the water of a certain locality is 10 mg per l ? Can we draw any conclusion regarding the time required if the magnesium content in water is 16 mg per l ?

10. The following table gives the chirps (X) per second of ground cricket and the temperatare (Y) at the moment in degrees Fahrenheit !

Chirps (X)	20	16	20	18	17	15
Temperature(Y)	88	72	94	84	81	75

(*a*) Does their exist linear correlation between frequency of chirps and temperature ?

(*b*) Fit a least square regression line to the above data for predicting chirps per second from various temperatures.

(*c*) Find the regression line to predict temperatures from the frequency of chirps.

(*d*) What is the expected temperature if a circket is making 19 chirps per second ?

(*e*) What is the expected frequency of chirps if the temperature is 90°F ?

11. The following table gives some data on learning (Y) and the duration (X) of training obtained from 6 students :

X *(duration)*	8	9	4	2	3	2
Y *(learning)*	9	7	7	5	2	2

(*a*) Find the equation of the line of regression of Y on X and plot it on the scatter diagram.

(*b*) Use ruler to read from the graph the predicted values (correct to one decimal place) of Y for each value of X given in the table.

(*c*) Use the above predictions to calculate the SEP of Y on X. Also calculate the same with the help of the formula and compare the results. How do you explain the slight difference in the two answers ?

(*d*) Determine the predicted values of Y for given value of X from the regression equation in (*a*). Compare with values obtained in (*b*). Which set of values do you think should be more accurate ?

12. Comment :

(*a*) A good example of bivariate data is provided by the ages of 10,000 girls in India and 10,000 boys in France.

(b) A low value of r indicates that there is hardly any type of relationship between the two variables under consideration.

(c) In case of perfect curvilinear relationship between X and Y, the value of r would be either 1 or -1.

(d) Pearson's r measures how closely the scatter diagram fits a curvilinear pattern.

(e) A negative value of r indicates that large values of X are accompanied by large values of Y.

(f) Evidence of a lurking variable implies a passive correlation.

(g) The closer the dots to the line of best fit, the lower the value of r.

(h) If X is the conditioning variable, Y the variable to be predicted and r the correlation coefficient, then when $X=X_0$, $Y=rY_0$.

(i) The least square line of Y on X gives the least value of the square of Y for a given value o X.

(j) The line of regression of Y on X is used to predict the values of X for given values of Y.

(k) The predicted values obtained from the line of regression of Y on X are, in a way, estimates of the mean of the various values of X corresponding given value of Y.

HISTORICAL NOTE

Correlation and regression techniques are used to study relationship between two (or more) variables. The study of *regression* in its current technical sense starts with Galton's work "Regression Toward Mediocrity in Hereditary Stature" published in 1885. With time however, the term regression has become standard for the statistical technique of predicting expected values of a variable from those of a related variable. It continues to play an important role in behavioural sciences in studies of individual differences, of role of heredity and environment, and of personnel selection and classification. However, today the techniques of correlation and regression are used in virtually all disciplines, even in business management and industry. In fact, any time several factors have influence over an event, a problem in correlation and regression is encountered.

□ □

JOHN VON NEUMANN (1903-1957)

John von Neumann, one of the greatest mathematicians of the twentieth century, was born in Budapest in 1903. From his early youth he exhibited remarkable talent in mathematics, physics, chemistry and engineering. After obtaining a degree in Chemical Engineering in 1923, he spent the early part of his career in Germany. In 1933 he was appointed professor of mathematics at the Institute for Advanced Study in Princeton, U.S.A.

von Neumann was one of the founders of the computer age. He played a central role in the design of some of the first U.S. electronic computers and in the development of programming techniques.

Computing

14·1. INTRODUCTION

What causes fear ? enquired a king of his wise minister. *Ignorance,* replied the minister, and demonstrated his point by sounding a shrill siren at the dead of night, frightening thereby all and sundry. *What causes awe* ? Let us ask and answer. To the extent awe is fear, it is caused by ignorance ; to the extent it is respect, it is caused by knowledge of the power and strength of someone or something. For example, all of us are awed by *computers.* Partly, this is due to our ignorance about computers, and partly due to our knowledge of the miracles that computers can and do perform.

Most of us perceive a computer as a calculating device. That is true, but only as far as it goes. As we shall see, it is not the whole story. However, we must admit that computers are the result of man's desire to *compute* or *calculate* with as little labour as possible. Man has performed a long and slow journey towards this goal starting from his fingers, and going on from there to sticks to notches to knots to abacus to gear driven machines to electro-mechanical devices to the modern-day electronic computers finally.

Most of us have some misconception or the other about computers, something worse than ignorance even. In this chapter we try to remove some of the myths surrounding computers by talking about them, both as a *tangible piece of machinery* as well as a *conceptual tool,* by hinting at some of its uses and abuses, and by concluding the chapter with the explanation of some terms that one hears about computers all around.

14·2. WHAT IS A COMPUTER

Logically, anything which *computes,* should be regarded as a computer. That, however, is not the common usage. Even the word-meaning *calculate* of 'compute' has a slight undertone in connection of computers. Technically, a computer today is understood to be an *electronic device that is capable of accepting data, processing these data automatically according to stored instructions, and then giving out the processed data in the form of information.* Too many unfamiliar and not-so-unfamiliar words ? Let us make friends with them.

Data. Data is the plural of the word *datum* meaning fact. However, it is generally used for any representation of quantities or characters which are capable of being assigned a meaning. Normally the word is used for raw information or facts which we wish to manipulate, put in a context, and present in a usable or meaningful form. For example, if we wish to find the sum of 3 and 4, then 3 and 4 are our *data*, *summing* them is *processing* them, and 7 is the *information* obtained. Or if we wish to prepare the glossary of a book, then various terms with their definitions are our *data*, *arranging* them *alphabetically* is the *processing* and *the ordered list* is the *information* obtained.

The data to be processed are generally called **input** and the obtained information is called the **output**.

Automatically. Refers to *without human intervention*. For example, when you stand on a weighing machine, out comes a ticket with your weight (and a message etc.). The weighing is *automatic*.

Stored instructions. In the weighing example, we had instructed the machine to weigh you, *i.e.*, we had built up a mechanism to weigh you and had already *stored* it inside the machine. The little digital watch on your hand starts beating seconds, or demonstrates the day and date just by the pushing of some buttons. It has been *instructed* to do so and the instruction has been *stored* in it.

It must be clear now that a computer is capable of processing data, *any data* that we are capable of feeding into it, according to *any instructions* we are capable of storing into it. Data can be numbers, instructions can be arithmetic operations; the output would then be the result of some calculations. Such operations today constitute only about 20% of the jobs that all existing computers do. Data could be 'characters' like English alphabets, like a list of words; instructions, as above, can be 'compare and order', in which case the output is an alphabetically arranged list. Now that is not numerical calculation; and believe it or not, but 80% of the total time computers are in operation, is devoted to such *non-numeric* tasks. That is why a computer is known as an *information processing tool* rather than a *calculating device*. Of course, it does process numbers like other bits of information. Computer technology is notorious for getting obsolete rather soon. As fate would have it, even the very name of the thing is already obsolete

Have you been wanting us to say that it is a machine? Yes, it is a machine. Are you wondering about its size and shape? Just as there are small fish and big fish, so there are small computers and big computers. In some sense, the small digital watch an your wrist is a computer. Then there are computers which are housed in large rooms. Size depends on what they are meant to do, and

when they were built. The time factor is vital not only to *size*, but also to *speed* and *cost*.

Size. The history of modern electronic computers started about fifty years ago. ENIAC, one of the earliest computers (brought in operation in 1946), was a huge machine. It used more than 18,000 valves, 1,500 relays ; about 500,000 connections were soldered to link these. It weighed thirty tons and occupied 15,000 square feet of space. It looked like a set of almirahs, and engineers had to use ladders in order to operate them. Today a computer with more power can be put on your study table. It more or less looks like a TV to which has been attached a sleek and beautiful typewriter kind of thing. There may be other small attachments.

Speed. Computers today are much faster than they were fifty years ago. ENIAC could do 5000 additions in one second. But 'second' is a large period in the terminology of computers today. We talk about *microsecond* (millionth of a second), *nanosecond* (billionth of a second), and *picosecond* (trillionth of a second). *A nanosecond is to a second what roughly a second is to 31 years. A picosecond is to a second what roughly a second is to 31,700 years.* Had our transport system become faster in the same proportion, we would have been travelling faster than light ! Computers today perform millions of operations per second ; some, even billions.

Cost. In the beginning, computers were a costly affair. No individual could dream of owning a computer. Today, most homes in the West have a computer. Prices have come down drastically. If prices of cars had fallen in the same proportion, you would be able to buy a Rolles Royce for about Rs. 15. Today you can buy a computer from Rs. 10,000.00 upwards.

14·3. BLUEPRINT OF A COMPUTER

You must have realized that every computer must have one or more components that accept data. These are known as *input devices*. There must be somewhere to keep the data. That space where data are kept, is known as *memory* or *storage devices*. The part looking after processing is known as the *arithmetic logic unit* or (*ALU*). The parts giving out processed information are known as *output devices*.

Since so much is going on inside the computer, there must be some part which controls all the activity. This part is known as the *control unit*. The control unit together with the memory and the arithmetic logic unit, is known as the *central processing unit* (*CPU*). This gives us the following map. The arrows indicate access. Thus an arrow from input devices to CPU indicate that these devices have an access to CPU. We shall now give the physical descriptions of these devices one by one.

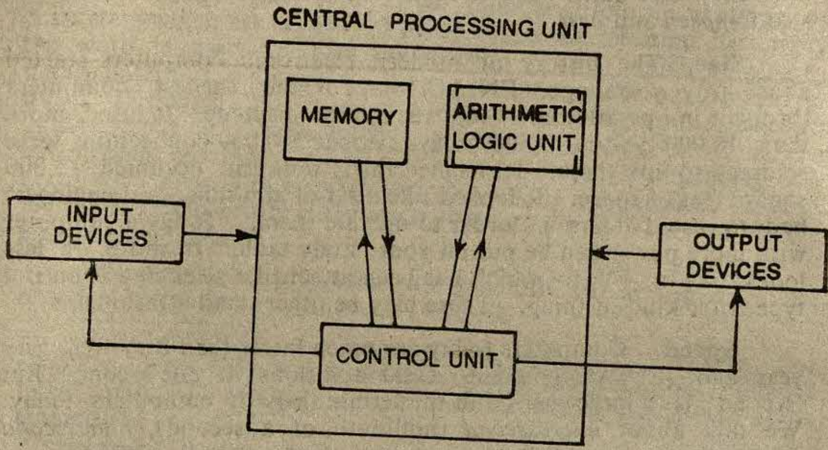


Fig. 14·1

14·3·1. Input

When you wish to telephone from a public telephone, you have to put in a coin of a certain denomination. To make your car go, you supply petrol, diesel, and water. A wood sawing machine requires logs of wood to be fed in. In what form does computer accept data or input ?

A computer is essentially an electronic device. It senses everything in terms of electric pulses. Different patterns of presence and absence of currents convey different things to a computer. The inside of a computer is a complex jungle of electronic circuits. A computer can do only those things which these circuits make possible. In 1938, an electrical engineer Claude Shannon showed that an electronic circuit could be made to perform arithmetical and logic operations by using the principles of Boolean algebra developed by the mathematician George Boole. Computers are distinguished by these circuits for one thing. The circuit inside your digital watch is different from that inside a computerized telephone. When you push a button on your watch, a certain circuit is activated and the watch displays day, date or does whatever other function like playing music etc. is meant to be bone through this circuit. Similarly, when you dial a number, find it engaged, then press a particular button, a certain circuit is activated ; the number gets on getting dialled automatically till it is connected and then there is a *ping* to tell you that you can talk.

At any time, either a current flows through a circuit or it does not. That makes computer a *two-state* or a *binary* machine. Now how do we go about representing these two states ? Clearly, we need two symbols to represent these two states. Any two symbols

would have done, but remember that initially computers were being devised by mathematicians to reduce the drudgery of calculations. It was, therefore, natural that 0 and 1 were chosen, being the first two of the ten symbols 0, 1, 2, ..., 9, given to the world by Hindus, and used all over for writing all possible numbers.

The task of building up numbers by means of the symbols 0 and 1 is rather simple. We have only to copy the fundamentals from the decimal system. Ten symbols lead to place values which are powers of ten ; two symbols mean powers of two. Thus for example, in the decimal system, the number 456 means 6 units, 5 tens and 4 hundreds. We can write it as under :

$$\frac{10^2 \quad 10^1 \quad 10^0(=1)}{4 \quad\ 5 \quad\ 6} = 4\times 10^2+5\times 10^1+6\times 10^0,$$

$$=400+50+6.$$

With two symbols 0 and 1, every number is written through 0 and 1 alone ; there is no 2 or 3 or 4 etc. Consider the number 1101, say. Writing it in powers of two, we get

$$\frac{2^3 \quad 2^2 \quad 2^1 \quad 2^0(=1)}{1 \quad 1 \quad 0 \quad 1} = 1\times 2^3+1\times 2^2+0\times 2^1+1\times 2^0,$$

$$=8+4+0+1=13.$$

Thus the number 1101 written in the number system having only two symbols 0 and 1, is equivalent to the number 13 in the decimal system. This new number system is known as the **binary number system** and its symbols 0 and 1 are known as the *binary digits* or simply as *bits*. Everything you can do in the decimal system, you can do in binary. Only binary arithmetic is much simpler though representations are longer.

People soon realized that just as they could *code* numbers like 13 by means of a string 1101 of one's and zero's, so they could *code* letters like A, B, C,......, *a*, *b*, *c*,......... and symbols like +, −, ;, ', ", etc. by special strings of one's and zero's. (A *code* is a set of rules that transforms data from one representation to another.) Thus A could be coded as 110001, the first two ones being supplied in a different manner in order to distinguish it from the binary number 110001 (=32+16+1 in decimal system).

The next step in this chain of codes was the real revolutionary step and it changed the face of computing by giving the computers an incredible speed. The idea of feeding the instructions just like data by coding them into strings of zero's and one's, was known as the concept of **stored instructions**. This concept is due to a mathematician John Von Neumann (1903-1957). This did away with the need of human intervention (like soldering the links) and computing became automatic. This means, we can now feed both data and the instructions as to what is to be done with these data, in the

beginning itself. All this we organize in terms of just two *symbols* 0 and 1. Later, one of these symbols is identified with the presence, and the other with the absenee of current. The components of the computer which help the CPU in sensing these current-signals are the input devices. Generally, one does not distinguish between the terms input and *input devices.* It is for the sake of clarity that we are using the first for data only. Similarly for output and output devices.

14·3·2. Input/Output Devices

If you are thinking that whatever you want to convey to the computer must first be coded by you into zeros and ones, then relax. Input devices are built to do the job for you. You only have to feed your data according to the device you choose. For example, if you choose the key-board, then you can type in all the numbers and alphabets just in the ordinary manner. Before we describe the various input/output (I/O) devices, we would like to impress upon you that though these are essential to a job being done by using computer, they are known as *peripheral* devices. By computer, one sometimes means its CPU. 'System' and 'Computer System' are synonymous terms used more often than the term 'computer' itself.

14·3·3. Punched Cards and Punched Card Reader

Punched cards were popular till a few years ago. The idea of punched card came into computing from Marie J. Jacquard, who used card-board cards with holes punched in them to control the design on cloth being woven by a handloom. Herman Hollerith used a card with holes punched in it to represent data. These data, in the form of holes, were recognized by his tabulator through a series of electro-mechanical brushes as the card was fed into it.

The modern-day punched card is a refinement of Hollerith's card. It contains 80 columns and 13 rows for the purpose of punching holes. Different patterns of holes represent different digits, alphabets and special characters like comma etc. The *card-reader* is the machine which accepts these cards and *reads* them either by its wire brushes or by a photo-electric process, generating impulses which complete specific circuits and thus convey data to the computer.

14·3·4. Punched Paper Tape (Reader)

Instead of individual cards, one uses a continuous roll of paper tape. The principle is the same as for cards and card readers.

14·3·5. Magnetic Tape

Instead of paper tape, one uses magnetic tape. Data are represented by patterns of magnetic spots rather than holes.

14·3 6. Magnetic Ink Reader

Numerals and other characters can be printed in magnetic ink. Computer can sense and record these different magnetic patterns. This device is being used in several banks of India as MICR (Magnetic Ink Character Reader).

14·3·7. Keyboard Devices

One can use a typewriter-like device, a keyboard (KBD), for feeding in data. Depressing a key sends an electronic pulse which completes a circuit. Data can also be transferred from KBD to a reel of magnetic tape or a magnetic disk (something like a record in a record-player) or a floppy disk (something like a magnetic disk but made of a plastic-like but elastic material, very light), and then to the computer. A visual display unit (VDU), which is very much like your TV screen, enables you to see the data being keyed in on the screen ; thus corrections can be made instantly.

Of the devices mentioned above, punched cards, punched paper tape, magnetic tape, KBD, and VDU are also used as an output device. Instead of a *reader*, you have a *puncher*. However, the main output device is the teleprinter or simply the printer.

We now describe one by one the three components that essentially comprise a computer system.

14 3·8. Memory (Storage Devices)

We have already met some of the storage devices such as magnetic tapes, magnetic disks and floppy disks where data is stored in the form of patterns of magnetic spots. You may be surprised to learn that one can store thousands and thousands of data items on a single disk and it only takes a microsecond or so to sense an item when needed. The memory available on these devices is known as **backing memory** or **secondary storage**. It is not housed within the CPU but can be accessed when needed. The main memory or the primary storage is housed within the CPU and most frequently is in the form of tiny metal rings known as *ferrite cores*. The direction of magnetization in these rings indicates a one or a zero.

Research is on for cheaper and faster (data can be *stored* and *retrieved* faster) memories. Some potential condidates are bubble memory, and laser disks. Gallium arsenide is another material that holds promise. A form of bio-chemical memory is also being used experimentally. It is faster than the present computers can manage.

The primary memory consists of a large number of cells, each capable of storing data, and each having a fixed location and a name (known as its **address**). $2^{10}(=1024)$ cells or memory locations constitute 1 kilobyte or 1K memory. The *capacity* of a computer is the amount of the core memory within the CPU, and is expressed in kilobytes (*i.e.*, K). Data, together with the instructions to tell the

computer what to do with it, are all stored in the primary memory for processing. Recall, both data and instructions are sent to CPU as strings of zeros and ones. Thus, cores can accommodate them.

14·3·9. Arithmetic Logic Unit (ALU)

Once data and instructions are there in the memory of the CPU, the ALU does the required processing and stores the result in the instructed memory location.

You have a vague feeling that computers are capable of performing mind-boggling mathematical feats in the twinkling of an

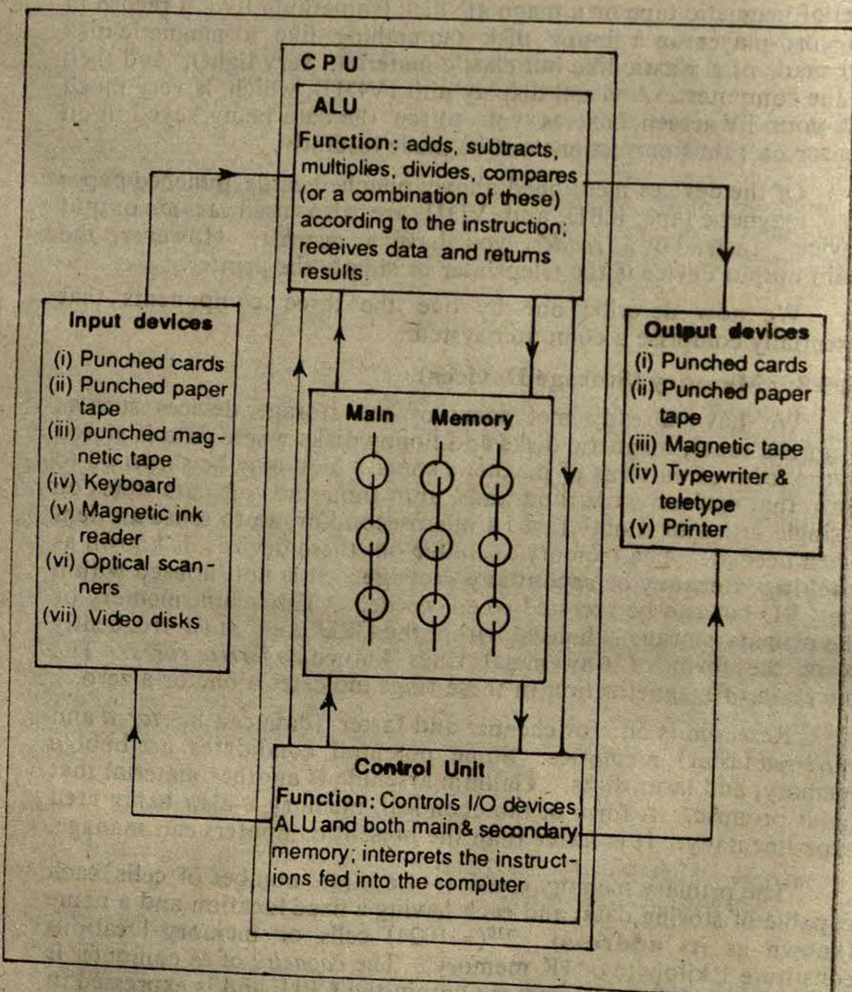


Fig. 14·2.

eye ; haven't you ? This is true ; but you must know that computers cannot really do any *complex* mathematical jobs. All they are capable of is the four arithmetic operations (in fact two ; multiplication is performed as a repeated addition, *e.g.*, 2×3 as $3+3$ and division is carried out as repeated subtraction) together with the logical operation *compare*. Only they are very fast and do not make mistakes. All the rest depends on the ingenuity of man to be able to specify jobs which can be done by combinations of these five simple operations. As you learn about computer applications, you would be surprised to know how much can be achieved through these jobs. Operations which require huge quantities of data to be processed, and jobs which are repetitive in nature are ideally suited to be done at a computer. Why computers have revolutionized science is due to the fact that computers can handle huge amount of data which it was humanly impossible to do earlier. Complexity of the problems has always been the lesser of the evils, and generally manageable. A very ordinary computer today can perform in one second all the calculations that a mathematician would have done during his life-time before the advent of computers.

14·3·10. Control Unit

This is the most important unit of the CPU. It is the control unit that looks after the processing, instructing the ALU to get appropriate data and perform on them the correct operations in the proper sequence. We can now enhance our blueprint as in Fig. 14·2.

EXERCISE 14 (*a*)

1. Identify true statements :
   (*a*) Computer is the modern name of a calculating device.
   (*b*) Computer is the machine version of the human brain.
   (*c*) A computer is nothing but a TV attached to a typewriter.
   (*d*) The real part of the computer is its peripheral devices.
   (*e*) Computers can count only upto 2.
   (*f*) Every device which can be used as an input device is an output device.
   (*g*) The memory which is part of CPU is known as the backing memory or the main memory.
   (*h*) Main memory consists of magnetic disks.
   (*i*) A user of a computer must know how to make electronic circuits.
   (*j*) Computers tell us how to solve complicated problems of mathematics.
   (*k*) Bit is a name for a 2-state device.
2. Write 2, 4, 8, 16, 32 in binary number system. Also write 7, 10, 15, 20 same way.

3. Convert the following binary numbers into their decimal equivalents :

   101, 1101, 10101, 1111, 1110.

4. Sum the following numbers in the binary system by using a process analogous to that for decimal system. Convert the numbers and sum into decimal and verify your answer.

(*i*)	(*ii*)	(*iii*)
1010	1110	1101
+101	+1001	+111
———	———	———

5. What is the biggest decimal number expressible by 2 bits ? 3 bits ? 4 bits ? What is the smallest decimal number which requires three bits for being expressed in the binary system ?

6. Describe the function of the following :

   (*i*) Input device (*ii*) ALU (*iii*) Memory (*iv*) Control Unit.

7. Write in full :

   VDU, ALU, CPU, I/O.

8. Distinguish data from information.

14·4. INSTRUCTING COMPUTERS

The working of a computer is quite often compared to the working of human brain. This analogy is given below :

Human brain	*Computer*
* five senses	input
* learning/commiting to memory	storing in memory
* recalling	retrieving data from memory
* communicating	output
* taking decisions and solving analytical problems	ALU

However, human brain has *intelligence* and can *think*. A computer is supposed to have no intelligence and its IQ is zero ; it cannot think, and does exactly as you tell it to do ; it is your most obedient servant. And to that extent, it is a detestable servant. It hardly ever does what you *want* it to do ; it does exactly what you *tell* it to do. There is such a gap between wanting and being able to tell ! This places a very big responsibility on the shoulders of those who want to get any work done by it. We have to be very careful as to what instructions we give it.

The set of instructions given to a computer for performing a job is known as a **program**. Be careful of the way you spell 'program'. The American spelling has become standard. There are various stages of a program. First of all you have to organize

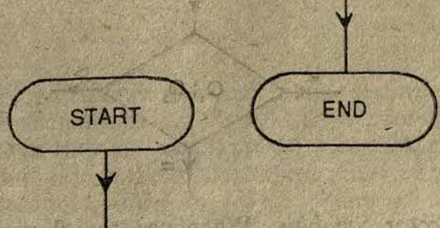
within your mind the plan of action in the minutest detail as your obedient servant has absolutely no initiative, and is completely devoid of any common sense of its own. This is where we come to realize computer as a conceptual tool. A *tool* is something with which to take things apart and re-assemble after you have achieved your objective. A concept is a generalized idea about a collection of things. A computer provides you an opportunity to play with ideas instead of things, organize them, test them.

Sometimes the job is simple and the plan can be made mentally. More often than not, the job that we wish to use the computer for, is complex. It is helpful to use paper and pencil. We analyze our problem and list out the various steps in the solution of our problem. These steps have to be taken in a certain sequence or order. Some of the steps simply involve reading/writing and input/output. Some of the steps are processing steps where we perform either an arithmetic or a logical operation (*i.e.*, compare). A logic operation normally requires a *decision* to be taken depending upon the result of the comparison. Quite often some steps may have to be *repeated*. *Flow charts* are a graphical device to represent the *procedure of the solution* or the *algorithm* consisting of the various steps outlined in the above fashion. We shall have more to say about algorithms later on.

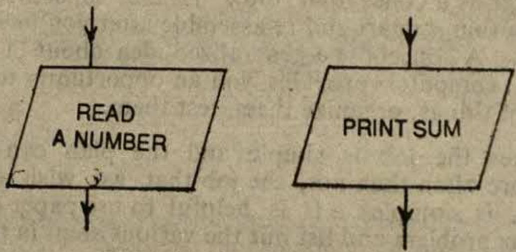
14·4·1. Flow-charts

The order in which we scan a printed page for reading is fixed in every language. In what order do you look at a picture ? Sometimes you look at it as a whole. Sometimes there are component parts. When the order or *sequence* in which they should be regarded is important, an arrow is drawn showing the order or the sequence. Flow-charts use this device for pointing the sequence of the steps in the program (*i.e.*, plan of action). The steps are written inside boxes of standard shapes depending upon their characteristics. We list below the main ones with their characteristics.

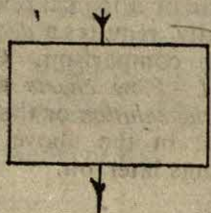
1. *Terminal Symbol.* An oval shaped box is used to start and end the program.



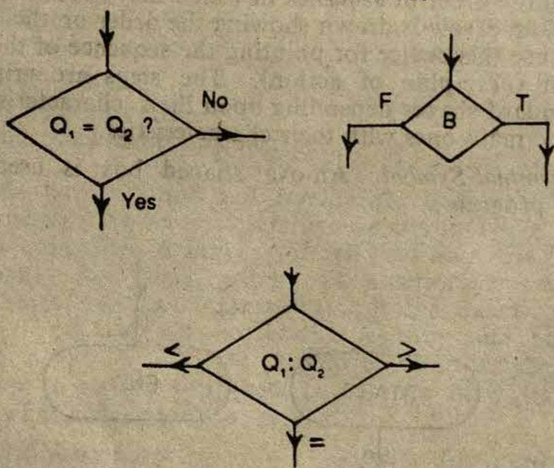
2. *Input/Output Symbol.* The I/O symbol is a parallelogram. It is used for steps intended to input data or get output from the computer.



3. *Processing Symbol.* Whenever a step is supposed to process data to produce output, a rectangular box is used.



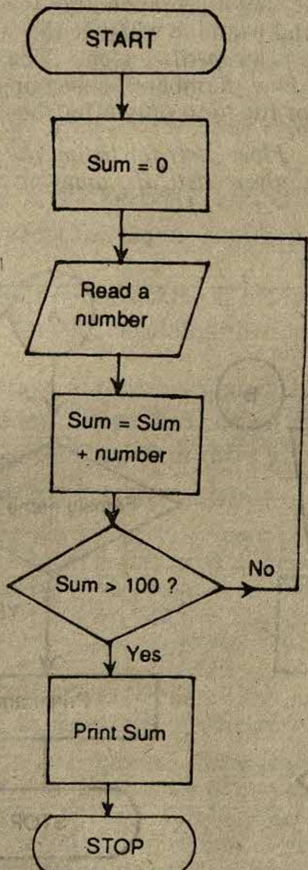
4. *Decision Symbol.* Whenever a decision is to be taken, we ask a question and accordingly plan our strategy. The question is put in a diamond and the various strategies are shown by two or three branches and the relevant path followed.



5. *Connector Symbol.* Whenever the flow-chart is carried over to next page or we do not want to draw too many flow lines

(arrows) to avoid messing the flow chart, we use a connector. It is generally a circular box with a letter inside it. Both the entry and the exit point are marked with the connector.

Illustration 1. *Flow-chart of the program for finding the sum of numbers from a given sequence of numbers until the sum exceeds 100.*



We start with an initial value zero for the quantity *sum*. Then a number is read. Equations like 'Sum=Sum+number' may sound confusing at first, but they are encountered frequently. The 'sum' on the right hand side is the initial values ; that on the left is the new value. For example, in the beginning, Sum=0. Suppose the first number read is 15. Then Sum=0+15=15. If the next number is 30, Sum=15+30=45 and so on. The equality is often called

the **assignment** and the symbol '$\leftarrow$' is also used instead of the equality symbol. Accordingly statements like 'Sum=Sum+number' are written as 'Sum$\leftarrow$Sum+number' and read as '*sum is* **assigned** *the value (current) sum plus the number* (read)', rather than as '*sum equal to sum plus number*'. Henceforth, we shall use the assignment symbol '$\leftarrow$' only. Every time a number is added, we find out whether or not the sum has exceeded 100. If the answer is 'Yes' we take the 'yes' branch and so we print the sum and stop. If the answer is 'no', we take the 'no' branch and read the next number, add it to the Sum and examine whether the sum has now exceeded 100, and so on. This is a *repetitive* step. The computer keeps on reading and adding new numbers, traversing the *loop* again and again. It comes out of the loop only after the sum crosses 100.

Illustration 2. *Flow-chart for listing the first name of a girl starting with P in a given list of names, assuming there is such a name.*

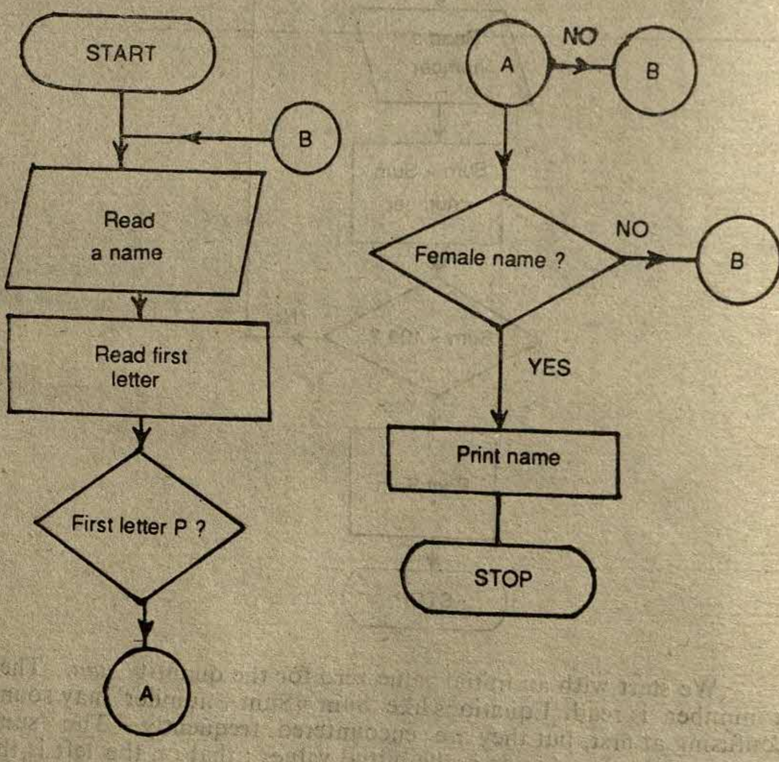


Fig. 14·3.

'Read a name' is input ; therefore, it is placed in an I/0 symbol. 'Read first letter' produces an output, viz., the first letter of the name read ; therefore, the processing symbol, rectangle, is used here. Below the first diamond, we encounter a connector A. Instead of continuing the flow-chart vertically down, we wish to save space, break the chart and put the pieces side by side. We find that there is a connector A on top of Fig. 14·3(*b*) ; so we continue from there. If the first letter is not P, we take the 'no' branch and encounter an exit-connector B. We locate the entry-connector B in Fig. 14·2(*a*) and continue along the flow-chart to read a name. The second exit-connector B in Fig. 14·3(*b*) also brings us to the entry-connector B in Fig. 14·3(*a*).

We shall have more practice on flow charts in Chapters 14 and 15.

EXERCISE 14 (*b*)

Prepare a flow-chart for the algorithm/procedure for each of the following :

1. Locking your door and checking it before leaving.
2. Addressing a letter and checking the address before posting it.
3. Finding the average of three numbers.
4. Finding the smallest integer n such that $50n > 1000$.
5. Printing (writing) the biggest of three given integers.

14·5. PROGRAMMING LANGUAGES

A computer program is a set of instructions which tells the computer what to do. A programmer is a person who prepares a program. A programming language is a language in which programs are written. By the way, the first acknowledged programmer happens to be a lady—Lady Ada Augusta, the lovely Countess of Lovelace and daughter of the famous poet Lord Byron. One programming language has been named ADA after her to acknowledge her contribution to the world of computing.

The only language a computer understands is an electronic impulse. The presence and absence of current can be represented by two symbols 0 and 1. Any number/character/instruction we want to input must be coded into strings of zeros and ones. That certainly is a tedious and time-consuming job. Moreover, it is error-prone. In the beginning, people talked to computers in this language only. The language was called the *machine language*, being the language of a machine.

Man is lazy by nature in that he finds means and ways to reduce his drudgery. People decided that computer itself be made to do the coding. They started writing programs in a *symbolic* language. They also wrote a program which could *assemble* (*i.e.*,

LADY ADA AUGUSTA

translate) this program into a machine language program. The translator program was known as the **assembler**, and was dependent on the characteristics of the computer on which it was to run. The symbolic language was accordingly called the *assembly language*. In order to get anything done from a computer, both the symbolic language program and the assembler (translator program) were loaded (fed) into the computer. The assembler translated the symbolic language program into a corresponding machine language program and the work was done.

As computers were called upon to do more and more complex jobs, longer and longer programs had to be written. Assembly language, though simpler than machine language, was still tedious and time consuming. People decided to write more compact programs, in languages closer to their own, and let computer do the translation. Of course, the translator program would have to be written, but only once and for all. These languages would be *portable* (you can carry programs written for one machine to another with slight modifications) because they would not be machine dependent. Machine part would be taken care of by the compiler (translator program). Such languages were called *higher level languages* and *procedure oriented* because they were written to carry out certain specific operations or procedures. This was needed so as to keep the vocabulary of the language as limited as possible. Writing a compiler is a tricky job ; it takes several men (expert

programmers) several months to write a compiler even when the vocabulary is rather limited. However, the rewards are well worth the labour.

There are numerous computer languages in use today but they are all far from our natural languages. Concentrated efforts are on to make computers understand natural languages but the main trouble is that natural languages are context oriented. Devoid of context, a phrase like *cut of sight out of mind* may be interpreted as *blind idiot*. The simple sentence *I see* may mean so many things. It might mean I am capable of vision, '...is visible to me', 'understanding has dawned on me' and it might be a sarcastic statement different from all the above.

14·6. USES AND ABUSES OF COMPUTERS

The biggest use of computers is in *data processing*, be this data numeric or non-numeric. For this reason, computers are labelled as our most patient, efficient and diligent electronic clerks. They look after our accounts, prepare pay-rolls and take care of all kinds of billing (telephone, power, water, computer-time and so on). They collect, sort and store our data, and present required information based on the same as and when required without delay.

Computers help the banks *in keeping accounts* and customer transactions. As already mentioned, some banks in big cities of India have already started using computerized cheques. People in West are considering the possibility of a money-less society, where all financial transactions would be carried out through computers without our need to exchange money physically.

Computers are being used in all kinds of *reservations*. Sitting in one city, it is possible to get a confirmed plane or railway reservation from one city to another provided these cities are linked by a computer network. You must have heard about *Indonet* which makes such facilities, and others like getting lots of information on your TV screen possible.

Computers are being used *in medicine* in a big way. CAT (Computer Assisted Tomography) scan has become a household word. The idea is to detect tumours and determine their exact size and position, a task often otherwise impossible. Computers keep track of moment-to-moment condition of critical patients, and automatically administer injections and medicine as and when required.

Computers are used *in town-planning*. By taking three-dimensional pictures, they help the town-planner to locate suitable sites. Moreover, they provide optimal plan for the network of roads, market-places and picnic-spots. Computers draw superb maps.

Computers are used for detecting minerals and oil below the surface of earth and water. They detect earth-quakes, cyclones and

fires in the interiors of jungles. They classify the types of flora and fauna, population trends, and *natural resources* region-wise. They control air pollution by giving warning at the prescribed level of pollution.

Computers are used *in factories* for actual production. They design and cut complicated tools and assembly parts. They keep control over stock and allocate optimal use for scarce resources. They decide what to produce, when to produce and how much.

Computers have changed the face of *modern offices*. There is no need for piles and piles of files. The data can be recorded on floppies and disks. Packages like word processor make life easy for the secretarial staff. Through such packages, letters can be edited as by a magic wand ; you can add or delete matter anywhere in a letter, the words shifting left or right as required ; if the same letter is to be posted to several people, it need not be re-typed again and again, and so on, so forth.

The use of computers *in meteorology* is known to everybody. The weather forecast that you see on your TV every evening has been possible only because there are computers which can process a huge amount of data in a very short time. The papers have been full of the item regarding purchase of super computers from the US for use in weather-forecasting. (Speed is one criterion by which to decide whether a computer can be called *super*). The photograph showing clouds etc. is not a photograph developed from a negative. It is rather a digital photograph drawn by a computer on the basis of some data.

The role of computers *in research* is well-recognized and now they are wedging their way *into education*. Because of computers it is possible for students to learn at their own pace without inhibition. The teacher need not undergo the drudgery of tutorial and marking.

Computer graphics have opened up new vistas *for architects, engineers and designers*. Designs can be turned over, taken apart, magnified and transformed in hundreds of ways by just changing data or by devices like the mouse, light pen, graphic tablet etc. It is possible to project changes on screen and examine the effect.

As robots, computers are taking over *hazardous jobs* also. They work in nuclear reactors. They rescue people from burning buildings and dive deep into seas where no man has ever ventured.

Simulation is the technique to *imitate* part or all the behaviour of a system on the VDU. Thus areas where direct experimentation is dangerous, costly, impractical or immoral, can also be studied, *e.g., flying a plane, warfare, forestry* etc.

The list of computer uses is long and must, therefore, remain incomplete. This is as good a stage to stop.

How computers are going to shape future is anybody's guess. What scares people is the fifth-generation computers which would demonstrate intelligent behaviour and would possess expert knowledge. In wrong hands, they can cause havocs. The other dangers are like the threat of an emotionless culture, and on a smaller scale, lack of privacy, monotonous jobs and unemployment. Computer-crime has already caused considerable concern. Bank accounts are being tampered with. Hackers are on the increase. (A **hacker** is someone who unlawfully gets access to private data and tampers with it. Hacking defense-data may be dangerous for the country.) Data stored on disks may get damaged for several reasons. This may lead to loss of a large number of records. The loss may be irreparable. Numerous such other doubts are raised. However, despite the *doubters*, computers are coming, and coming with a (big) bang. Study of computers or the *Computer Science* is really and truly the *Future Science*.

14·7. ALGORITHMS

You have seen a great variety of things that a computer can do. Do you know what it cannot do ? It cannot do precisely the things we are unable to explain to it in the greatest detail. Its I.Q., as you know, is zero. It solves our problems only when we are able to tell it precisely what to do so that the result may turn out to be the solution we want. Thus in order to get a solution from a computer, we have not only to understand the whole procedure which would bring about the solution, we have to explicity express it also. This *procedure* or *process* or *technique* or *method* which can be regarded as a *sequence of actions taken according to some specified rules on some specified objects* to produce the desired result is roughtly what we call an algorithm. (You didn't by chance read the word as *logarithm* ? Notice that either can be obtained from the other by a mere rearrangement of the first four letters, but their meanings are a world apart) In order that a procedure may qualify as an algorithm, it must have the following five properties :

1. Finiteness. Consider the task of writing p/q as a decimal, where p, q are positive integers. The procedure involved is quite simple. We divide p by q and the moment we get the final remainder R, we add a decimal point to the quotient, multiply R by 10 and continue the process of division etc. This is a valid procedure for the task in hand. Now suppose $p=1$ and $q=3$. How long shall we go on ? So long as our patience lasts us. How long would computer carry on ? For ever (subject to the physical failure of the machine !). Thus a procedure which does not terminate after a finite number of steps is no good in the context of computers.

Here, we require that our procedure or algorithm *must always terminate after a finite number of steps.*

2. **Definiteness.** Often in life, we have to get things done from others. We describe to them very clearly what to do and how to do it; yet the desired result is not obtained. You send your younger brother to bring bread. He brings bread but it is not the bread you eat everyday; it is a different brand (you particularly *dislike*). Woe on him; doesn't he see which brand is used in the house? However, this is precisely the trouble. We *see* so many things that the other party does not. A 'bread' to us may be the 'brand of bread we eat everyday' but it is just 'bread', any bread to the other. That is what the difference in *'wanting' and 'being able to tell what you want', is.* You have to express it in the most *unambiguous* and *definite* terms. Unfortunately, *unambiguous* is a very *ambiguous* term. What is *unambiguous* to Sheela may be quite *ambiguous* to Leela. Thus our procedure must be described in most *definite* terms. *All the steps, all the actions, must be described in terms which have a certain fixed meaning.* This is where computer languages enter. In order that a procedure may qualify as an algorithm, it must be *definite*; in the frame of reference we are working in, it should do precisely what it is meant to do.

3. **Input**: A procedure for which there is no data to be processed is not an algorithm. Suppose we describe the procedure for preparing the monthly salary of employees in a factory. If now data are not available regarding any employee, what will be the result? Nothing. It is futile to talk about such procedures. Thus an algorithm *must have at least one input.* Sometimes zero inputs are allowed by some people.

4. **Output**: Of necessity, an algorithm must produce at least one output related to the input(s).

5. **Effectiveness.** Last, but not the least, an algorithm is an effective procedure. By effective we mean that *it is a procedure with all its steps being such that, in principle, they can be carried out by a person using paper and pencil in a finite amount of time.* For example, arithmetic operations of integers are effective procedures.

To sum up, an algorithm is a procedure consisting of some rules and actions which convey a definite meaning without any ambiguity in a given context. The actions when taken on given inputs in the specified sequence should solve one of a class of problems. The solution related to the given inputs should be obtainable in a finite number of steps. We can now define an algorithm formally as follows:

Definition. *An effective procedure which consists of a set of definite rules (steps), specifies a sequence of actions that provides the solution (output) to a given class of problems on the basis of*

one or more inputs in a finite number of steps, is known as an **algorithm.**

A better term to use in place of '*finite*', could be '*reasonable*'. 'Finite' may be disastrously big for all we care. Similarly, an algorithms may require so much storage that we may not care to use it at all. Thus with every algorithm there is associated a measure of *goodness*. The study of questions relating to this aspect of algorithms is one of the most challenging areas of computer science, but outside the scope of this book. We shall be content to consider some basic features of algorithm and ways to represent algorithms.

14·7·1. Representation of algorithms

An algorithm as we have seen is a chain of thoughts meant to solve a class of similar problems. We *can* represent it in many ways. Some popular means are listed below :

1. *Natural languages* : We can give the description in a natural language such as English, German, French and so on, and so forth. For example, an algorithm to solve a linear equation $ax+b=0$, $a\neq0$, could be represented in the English language as follows :

Compute the negative of the constant term. Divide it by the coefficient of the variable. The number so obtained is the required solution. Such representations have many drawbacks. If an equation is $3x+4=0$, what is the number obtained on dividing $=-4$ by 3 ? Thus it is quite vague to anyone who does not already know how to solve a linear equation. Other than vagueness, the description may be too long ; the logic may be difficult to follow and certainly, the computer cannot process it.

2. *Decision tables* : A decision table is a table which lists all the possible conditions (like in case of $ax+b=0$ above, (*i*) $a=0$, $b=0$, (*ii*) $a\neq0$, $b=0$, ...) and the actions to be taken in all the cases (like (*i*) every real number is a root, (*ii*) 0 is the unique root, ...). Decision tables are suitable for complex decision processes but highly unsuitable in case of simple decision situations. This is so because the logic of simple situations is otherwise transparent. In case of complex situations, we need to systematically record all possible conditions. Computer languages are available for translating decision tables into programming languages.

3. *Flow charts* : Flow-charts provide a very effective visual presentation of algorithms. The logic structure of the algorithm becomes very clear in this mode. It is possible to use the same flow-chart for representing the algorithm in any computer language whatever. It is possible to construct the algorithm in modules (independent sub-problems) bit by bit. To understand a flow-chart representation, one needs very little instruction, whereas, to under-

stand an algorithm written in a computer language, one has to first learn the language. Another great advantage of flow-charts is observed at the time when we have to modify our algorithm either due to the detection of a logic flow or due to a slight change in the problem.

4. *Computer languages* : So far as the implementation of the algorithm (running it on a computer to get the solution to a problem) is concerned, this form of representation is the most ideal. This is the ultimate goal, but it is a very subtle job generally. Before representing an algorithm in a computer language, it is generally convenient to understand the logic through a flow-chart or pseudo code (explained below). To present algorithms in a computer language, we must first learn and master it. When represented in a computer language, an algorithm is known as a **computer program.**

5. *Pseudo-codes* : *Pseudo* stand for sham, false, *deceptively resembling*. Coding, among its other connotations, means the process of writing an algorithm by means of a computer language. Thus a pseudo-code is not exactly a computer language, but somewhat resembles it. Pseudo-codes have gained in popularity over flow-charts because they are more (*i*) compact, (*ii*) like natural languages, and (*iii*) like computer languages than are flow-charts. Pseudo-Codes consist of expressions like "IF......THEN.........ELSE.........", "WHILE.........DO.........", "REPEAT UNTIL.........", and so on, so forth. We shall soon discuss them in detail.

14·7·2. Problem-analysis and Construction of Algorithms

Since an algorithm is meant to solve a class of problems, we must first understand the problem. To do this, we must kind of take it apart and see what makes it tick. This process of analyzing the problem is bound to suggest a method of solution in general. Once we can *see* a method of solution, we may try to list the important steps in the solution, and then elaborate each step. We can then test whether it works. More often than not, we may have to modify it. Thus there are two major tasks at hand-*problem analysis* (or understanding the problem) and *devising an algorithm* (or planning the solution).

Problem-Analysis. The first stage in the analysis of a problem is *defining the problem*. More often than not, problems are not stated so very directly. A little thought is needed to fix what exactly it is that we are asked to do. For example, suppose the problem is to find the best way to reach a friend's house. Looks innocent enough ! Yet there are two traps. First, what does *way* mean ? Does it mean the *route* to be followed, or does it mean the *manner* (on foot, on a cycle, on a bus,...........) in which to reach there ? Suppose it is the route. What does *best* mean now ? Does it mean the *shortest* route, or does it mean the *least crowded* route ?

Thus a *more precise statement of the problem is needed before* we think of a method of solution. What is the fun in solving a *wrong* problem ? Understanding the vocabulary of the problem is very important. Consider similarly the problem : Show that there are no more than 30 primes less than 100. A casual student may start testing every number less than 100 for being a prime, make the actual list, count the number in the list and show that it is less than 30. A little thought immediately tells us than we are not really interested in finding the primes less than .00 as such, and our labour may be reduced quite a bit. Thus *a more complex problem than that given, might be solved* and effort wasted, if we do not precisely define the problem. Another danger involved is that only a part of the problem may be solved.

The second stage in problem-analysis is the *identification of problem inputs and variables.* Sometimes this is easy to do at the problem definition stage itself. The inputs consist of the data to be processed by the algorithm, etc.

The third stage in problem-analysis is the *statement of the goal(s) to be achieved.* This generally involves a precise statement of the output and the format in which the output is required. In case of scientific data, this is generally a straight forward affair Not so with other applications however. Business reports are either too crammed up with details (and hence not read) or too sketchy (and hence unsatisfactory).

The next stage is the most challenging of all. We must set ourselves how the difficult task of *collecting information relative to our problem and the rules which may be used.* There is no set procedure to do this and every problem may be a class by itself in this respect. That is why we hear the often quoted phrase : *Problem-solving is an art.*" Like any art, it must be practised in order to gain skill. The best teacher here is *experience.* A general way to proceed is to ask ourselves some such questions :

1. Do I understand each and every term in the statement of the problem ?
2. What information has been given ?
3. What do I wish to find out ?
4. What data are to be used ?
5. What do I do to these data in order to achieve my goal ?
6. What additional information, not already given in the statement of the problem may be useful ?
7. In what format shall I output the solution ?

The last stage in the analysis of the problem is to ask ourselves *whether the problem needs to be solved by means of a computer* at all. Use of a computer is advised under the presence of one or more of the following conditions :

1. Too much computation involved.
2. A problem to be solved again and again for diffesent inputs.
3. The solution requires the repetition of certain specified steps a large number of times.
4. No set method of solution is known in advance, and experience must be gained through a large number of experimental inputs in order to guess the form of solution.

Once we have analyzed the problem, a method of solution can be devised. It consists of a sequence of actions to be taken in a specified order on the given input and the successive intermediate outputs respectively, and produces ultimately the solution. The various states and actions can be explicity stated as the problem-space of the given problem. Consider for example, our solution of the linear equation $ax+b=0$, $a\neq 0$. Here, the input (a, b) in this order specifies the coefficient of x and the constant term. We may call it the initial state x_0. The procedure consists of taking the negative of b and outputting $-b/a$. This involves two actions A_1 and A_2. The first action A_1 when taken on the initial stage $x_0=(a, b)$ produces an intermediate output $(a, -b)$. Thus we may say

$$X_1=A_1(x_0)=A_1((a, b))=(a, -b).$$

The second action A_2 taken on $x_1=(a, -b)$ produces $-b/a$, *i.e.*, it produces (second co-ordinate)/(first co-ordinate). Hence we may say $x_2=A_2(x_1)=A_2((a, -b))=-b/a$, or that $A_2((p, q))=q/p$. Thus our problem space P is given as

$$P=(\{x_0, x_1, x_2\}, \{A_1, A_2\}).$$

Here (*i*) x_0, x_1, x_2 are the three states of the problem. x_0 is the initial and x_2 the final state. (*ii*) A_1 and A_2 are the two actions. Input for the first action A_1 is the initial (first) state x_0; the output from A_1 is the second state x_1. The input for the second action A_2 is the second state x_1 and the output from it is third state x_2 which is the goal state in this case. The problem is solved in two steps.

The problem we chose was rather a simple one. Every state warranted a unique action. This in general, does not happen. An algorithm is meant to solve a class of problems and different inputs for different problems may behave differently. The algorithm must cover are possible behaviours. Thus depending upon the various conditions a state may satisfy, different actions may be warranted. For example, if $D=b^2-4ac$ is the discriminant of a quadratic equation $ax^2+bx+c=0$, then $D<0$ gives complex roots and $D\geqslant 0$ gives real roots. If we are working in the frame-work of real numbers, in the first case our output would be "no real roots exists", and in the second case we shall have to compute the two real roots.

It may be convenient to show all this by means of a *graph* consisting of points called **vertices** or **nodes** and lines (straight or curved) called **edges**. The initial point of each edge is the state on which the action corresponding to this edge takes place ; the terminal point is the state this action produces. Each vertex represents a state and each edge an action. Thus the graph of the example we considered above consists of three vertices (states) and two edges (actions) as shown in Fig. 14·4. The solution consists of the path

$$x_0\ A_1\ x_1\ A_2\ x_2.$$

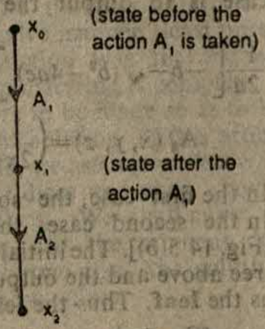
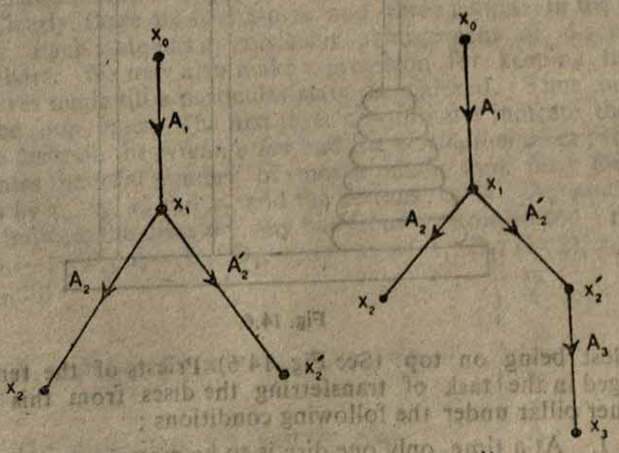


Fig. 14.4.

Coming back to the quadratic equation $ax^2+bx+c=0, a\neq 0$, the initial state is the input state $x_0=(a, b, c)$. As a first step in the solution, we may calculate the dlscriminant $D=b^2-ac$. Hence let

$$A_1\ (x_0)=(a, b, b^2-4ac)=x_1.$$

We cannot afford to lose a and b yet because they will be needed. If $b^2-4ac<0$, we wish to output "No real roots exist." Otherwise, we wish to compute $\frac{1}{2a}[-b\pm\sqrt{(b^2-4ac)}]$. Thus our graph at this stage must branch into two edges corresponding to the two actions. This gives us Fig. 14·5 (*a*), where $A_2(x_1)=$"No real roots exist".$=x_2$, and $A_2'(x_1)=(a, b, \sqrt{(b^2-4ac)})=x_2'$. The next action A_3 in the latter



(*a*) (*b*)

Fig. 14.5.

case is to output the two roots as $\frac{1}{2a}\left[-b+\sqrt{(b^2-4ac)}\right]$ and $\frac{1}{2a}\left[-b-\sqrt{(b^2-4ac)}\right]$. Hence we may define A_3 as

$$A_3(x, y, z)=\left(\frac{1}{2x}(-y+z), \frac{1}{2x}(-y-z)\right)=x_3.$$

In the first case, the solution consists of the path $x_0\ A_1\ x_1\ A_2\ x_2$. In the second case, the solution is the path $x_0\ A_1 x_1\ A_2' x_2'\ A_3 x_3$ [Fig. 14·5(b)]. The initial or the input state x_0 is called the **root** of the tree above and the output state x_2 or x_3 as the case may be, is known as the **leaf**. Thus the leaf denotes the output.

Wondering at the terminology ? Roots are below and the leaves above, right ? Our trees are *inverted* if you do not mind. This gives us the freedom to enhance them as we build up the solution bit by bit. Sometimes trees are also drawn left to right as we have done earlier in the text. It is more a matter of convenience than convention.

Illustration 3. (Brahma's Tower or Tower of Hanoi). As the story goes, in the Hindu Temple of Varanasi, there used to be three diamond pillars and sixty-four golden discs, all of different radii, placed on one of the pillars in decreasing order of radii, the

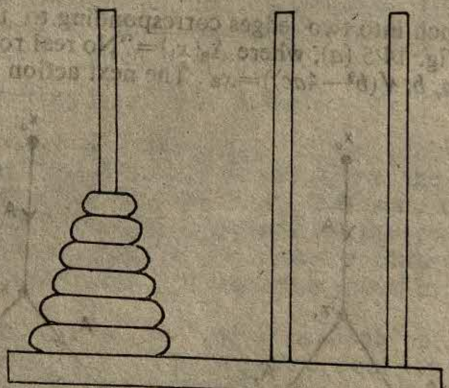


Fig. 14.6.

smallest being on top (See Fig. 14·6). Priests of the temple were engaged in the task of transferring the discs from this pillar to another pillar under the following conditions :

1. At a time, only one disc is to be moved.
2. No disc can ever be kept over a disc of smaller radius.

Thus at all times, the discs on every pillar are in decreasing order of radii from bottom to top. The transferring work was carried out incessantly in the most effective manner (least possible number of moves to be made). The prediction was that the world would come to an end, the moment the transferring is complete.

Let us see how the various moves are to be made so as to keep the number of transfers a minimum. The solution of this problem uses a *recursion* technique. In other words, we assume we know how to make the transfer when there are $n-1$ discs and show how to solve the problem when there are n discs. We shall cover this aspect of the problem in the next section. At the moment let us cover the initial states. If there were only one disc, the solution is straight-forward. Let us assume there are two discs.

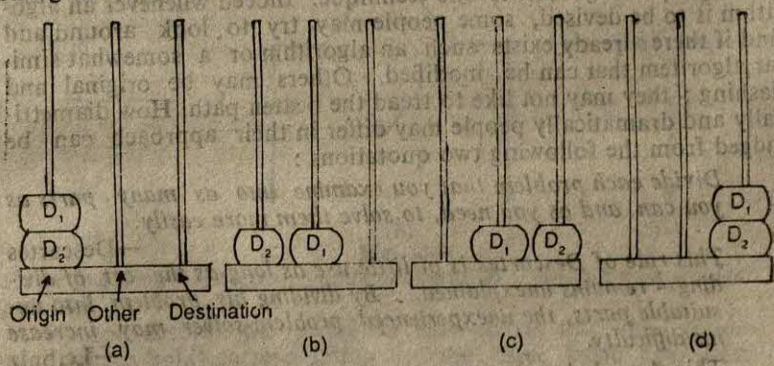


Fig. 14.7.

The solution is easy and shown pictorially in Fig. 14·7 (*a*) to (*d*). Clearly, there are four states and three actions in the problem space. Each state has a particular arrangement of discs on the three bars. We may also make a provision for keeping the count of moves made till a particular state is achieved. Thus our states may be four-tuples. The first three co-ordinates indicate the status of the discs on the *origin*, *other* and the *destination* discs ; the fourth indicates the total number of moves made thus far. Denote the states by x_0, x_1, x_2 and x_3, and the actions by A_1, A_2 and A_3. We shall indicate the discs on any bar from bottom to top. Thus $x_0=(D_2D_1, -, -, 0)$, $x_1=(D_2, D_1, -, 1)$, $x_2=(-, D_1, D_2, 2)$, $x_3=(-, -, D_2D_1, 3)$.

$$A_1\,(x_0)=x_1,\ A_2\,(x_1)=x_2,\ A_3\,(x_2)=x_3.$$

The initial state is x_0 and the goal state is x_3.

$$\begin{array}{c} x_0 \\ \downarrow A_1 \\ x_1 \\ \uparrow A_2 \\ x_2 \\ \uparrow A_3 \\ x_3 \end{array}$$

14·7·3. Construction of Algorithms

As has been emphasized earlier, there is no universal technique to generate algorithms. It is an art which draws heavily, both from the nature of the problem to be solved and on the way of working of the person devising the algorithm. For example, for solving a quadratic, someone may prefer to use the formula but someone may like to break it into two linear factors and then solve the two linear equations. Some people may prefer to have a very rough outline of the whole procedure and then refine the various components. Others may prefer to work on the components first and may wish to put the piece together latter. Sometimes the earlier approach and sometimes the later may be better ; yet out of habit some people may stick to one technique. Indeed whenever an algorithm is to be devised, some people may try to look around and find if there already exists such an algorithm or a somewhat similar algoritem that can be modified. Others may be original and dashing ; they may not like to tread the beaten path. How diametrically and dramatically people may differ in their approach can be judged from the following two quotations :

Divide each problem that you examine into as many parts as you can, and as you need, to solve them more easily.

—Descartes

This rule of Descartes is of little use as long as the art of dividing··· remains unexplained.... By dividing his problem into unsuitable parts, the unexperienced problem-solver may increase his difficulty.

—Leibniz

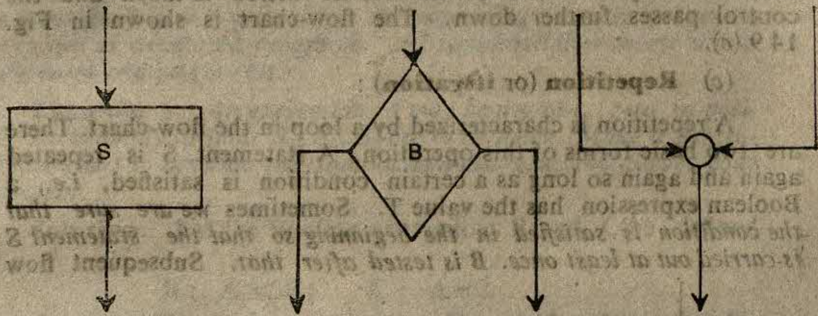
This then being the state of affairs, we shall not devote much time on this phase except for pointing out a *style* (viz., structured programming) of problem-solving which is gaining popularity amongst the computer-scientists. We shall also dwell on the main control structures of this style and express them through flow-charts and pseudo-codes.

14·7·4. Structured Programming

We have seen that sometimes a problem can be solved by carrying out a number of actions in a certain *sequence* one after the other like the solution of a linear equation earlier. Sometimes, we have to take a decision or to make a *selection* depending upon the particular condition a state may satisfy, like in our solution of the quadratic earlier. Another frequently occurring phenomenon that is encountered while solving problems is that of *repetition*. We may have to repeat a process again and again in order to achieve our goals like "reading a number" and "sum=sum+number" in illustration 1 on page 764. All this must be a matter of common experience to you, but here is something which might come as a surprise to you. Any reasonable problem which has a solution, can be solved by some combination or the other of the above three constructs or operations. Two Italian mathematicians Böhm and Jacopini proved that the flowchart for any procedure can be bulit

by a suitable combination of the flow-charts for these three basic constructs.

A flow-chart is a directed graph whose vertices are the various boxes, the flow-lines being the edges. There are three basic types of vertices—*function vertex*, *Boolean vertex* and *collecting vertex* shown in Fig. 14·8 (*a*), (*b*) and (*c*) respectively. Earlier, in our flow-charts, collection vertices where two incoming branches merge and pass control to a single branch, were not shown explicitly. A collection vertex is shown by a small circle instead of a box.



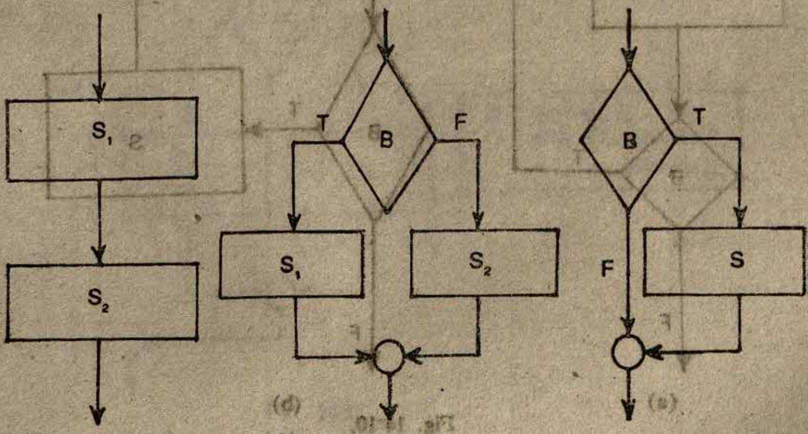
(a) (b) (c)
Fig. 14·8.

S above is called a (program) *statement* and B is known as a *Boolean expression.*

We now describe the primitive flow-charts and the pseudo-codes (written bold) corresponding to the three basic operations sequence, selection and repetition.

(*a*) **Sequence (or composition)** :

The pseudo-code is "*do* S1 ; S2". *Do* may be omitted. See Fig. 14 9 (*a*) for the flow-chart.



(a) (b) (c)
Fig. 14·9.

(*b*) **Selection (or alternation)** :

The pseudo-code is "*if* B *then* S_1 *else* S_2". This means if the Boolean expression B is true (T), then the program passes control to S_1 and the statement S_1 is carried out. Otherwise [B is false (F)] control goes to the other branch and S_2 is carried out. The flow-chart is given in Fig. 14·9(*b*). A special case of "if B then S_1 else S_2" often encountered is "if B then S". In the general case, one of the two actions S_1 and S_2 is bound to be taken. In this case, the action S is taken only if B is true ; otherwise no action is taken and the control passes further down. The flow-chart is shown in Fig. 14 9 (*c*).

(*c*) **Repetition** (or **iteration**) :

A repetition is characterized by a loop in the flow-chart. There are two basic forms of this operation. A statement S is repeated again and again so long as a certain condition is satisfied, *i.e.*, a Boolean expression has the value T. Sometimes *we are sure that the condition is satisfied in the beginning so that the statement S is carried out at least once. B is tested after that.* Subsequent flow

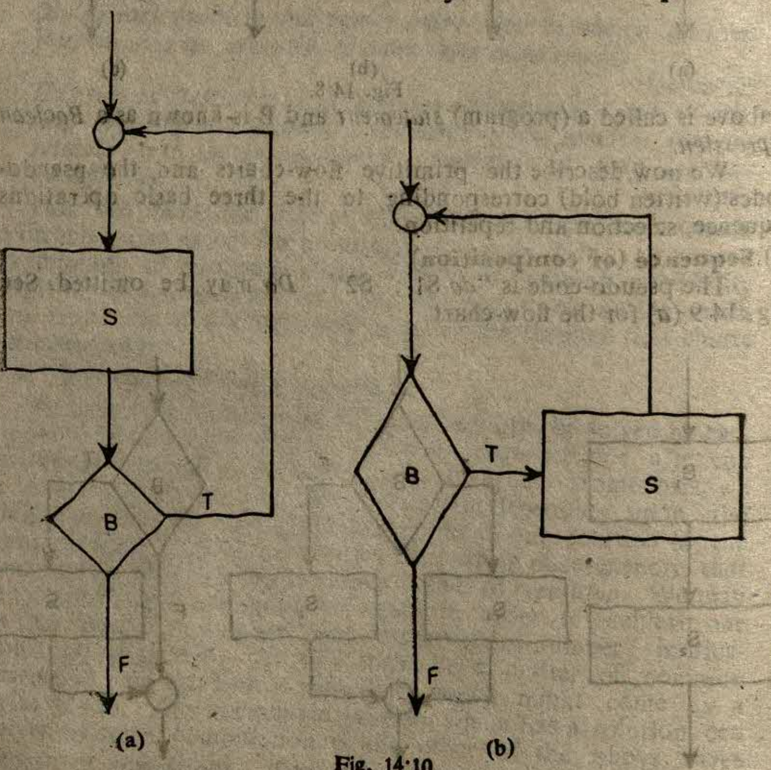


Fig. 14·10.

of the program is determined by the result of the test. This situation is shown in Fig. 14·10(*a*) and the pseudo-code for this is "*do* S *while* B".

Sometimes the *condition B is tested right in the beginning and* the statement S gets control only when B is true. Quite likely, B is found false the first time itself. Then *S need not be carried out even once.* Such a loop or iteration is expressed as "while B do S". The flow-chart is shown in Fig. 14·10 (*b*).

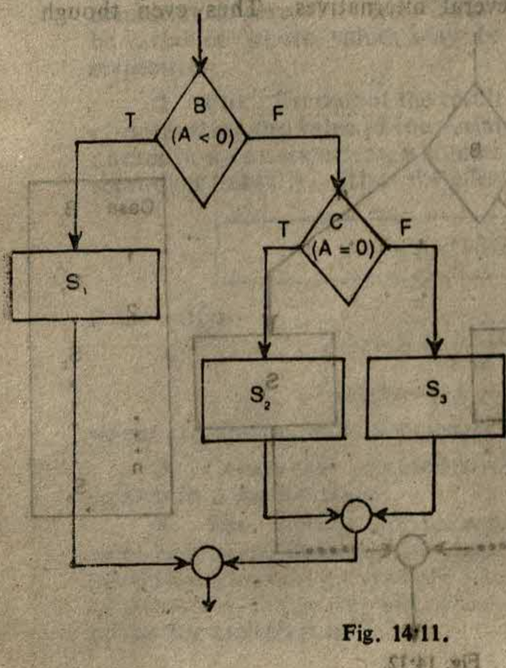
Any flow-chart based on these primitive flow-charts is known as a structured flow-chart and the corresponding computer programs are known as structured programs. *All structured flow-charts have a single entry and single exit.*

Example 1. *An expression A may be negative, zero or positive. Accordingly, actions S_1, S_2, S_3 are to be taken. Make a structured flow-chart to describe the situation and also give the pseudo-code equivalent.*

Solution. Let the Boolean expressions B and C be defined as

$$B:\ A<0; \qquad C:\ A=0.$$

Then the required flow-chart is given in Fig. 14·11 and the pseudo-code equivalent in the adjoining frame.



If B then do S_1
else If C then do S_2
else do S_3

Fig. 14·11.

Notice that here we combined two **if-then-else** statements to consider a condition which had three alternatives. When such is the case, the successive **if-then-else** statements are indented ; '**then**' and '**else**' are written one-on-top-of-the-other fashion. In terms of flow-charts, such a situation generates *a loop within a loop* (*i.e.* nested loops). A condition with more than three alternatives can be covered by using more diamonds (or more nested loops). This means that the innermost **else** and **then** form one statement ; then the next innermost else and then are to be combined and so on, so forth. In this respect, the various **elses** and **thens** behave like nested-brackets. The associated **else-then** should be written one-on-top-of-the-other fashion to facilitate understanding. See, *e.g.*, the following nested statements :

```
If B then if C then do S1
                 else do S2 while D
      else do S3 ; S4
```

You must have realized that our flow-charts can be very messy and the pseudo-code equally so, if we are faced with the task of handling a condition with several alternatives. Thus even though

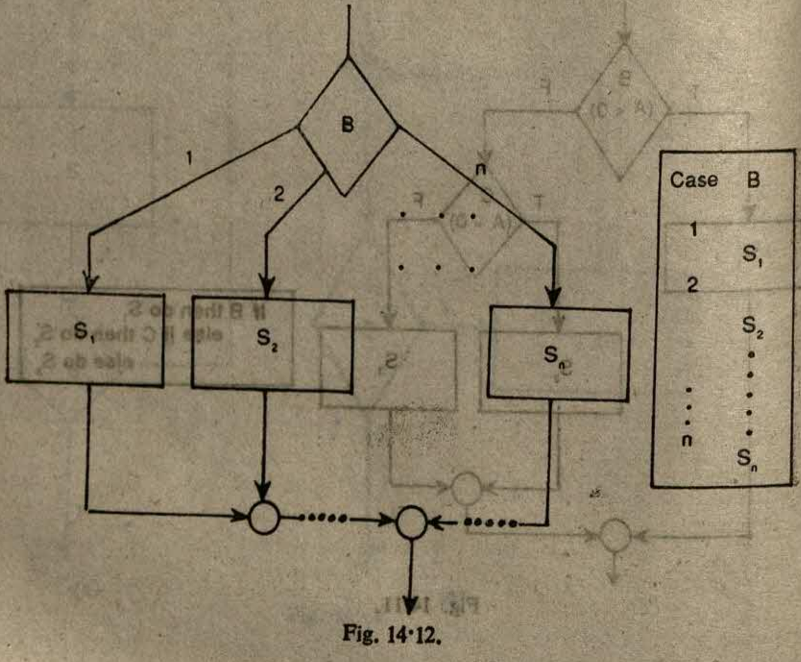


Fig. 14·12.

Böhm-Jacopini's constructs are *sufficient*, they are not the most *convenient* or *natural* to use. We shall consider two easy and convenient alternatives, one to avoid the nested loops and one to facilitate the repetition operation.

The **case** *construct* : The case construct is a generalization of the **if-then-else** construct. The *cases* are the various alternatives. Instead of nesting the loops by two branches going out of each Boolean vertex, we may send out several branches one corresponding to each **case** and then collect all of them by a number of collection vertices to send out a single branch. The flow-chart for the **case** construct is shown in Fig. 14·12, where it is assumed that there are n cases in all, the jth case necessitating the implementation of the statement S_j. The pseudo-code is given in the adjoining frame.

14 7·5. Examples and More Pseudo-code

We shall now solve some examples. The following terms would be used as part of the pseudo-code :

1. **Get.** To fetch values of variables. Thus "**get** x" has the effect of making the value of the variable x available for the rest of the program.

Remark. Variables in the context of computers need not be denoted by single letters like x, y etc. Generally one prefers to use mnemonics (the first *m* silent ; a term suggestive of the meaning) consisting of several letters. Thus TIME, NAME, INT, SPD may be variables whose values may be times, names, interests, speeds respectively.

2. **Put** : To output the result. Thus "**put** y" has the effect of outputting the value of the variable y. Whenever we use double quotes in a **put** statement, whatever is enclosed within the quotes, is reproduced exactly. Thus the effect of the **put** statement

```
put "HCF is" REMA
```

is the output

```
HCF is 11
```

where 11 happens to be the current value of REMA.

3. **Comment.** Is used to write a comment in a program written in a pseudo-code.

4. **For.** This is used for *repetition* when we know in advance as to how many iterations are required. It avoids the testing of condition for getting out of the loop. After the specified number of iterations, the getting out of the loop is automatic. The format of the **for** construct is

```
for I = I0 to In by h do S
```

where (*i*) I is a numeric variable name or identifier,

(*ii*) I_0 is the initial value of I,

(*iii*) I_n is the *test* value of I, and

(*iv*) *h* is the *step* or *increment* (or decrement) by which the value of I is to be changed each time.

Thus "**for** I=0 **to** 100 **by** 2
do SUM←SUM+I"

has the effect of adding the numbers 0, 2, 4, ···, 100 to the value of the identifier SUM.

Remarks. 1. The process terminates as soon as the value of I becomes greater than the *test value*. Action is taken only till the value of I is less than or equal to the test value, assuming that the increment step is positive.

2. When the step is 1, it is generally not mentioned. Thus "**for** I=1 **to** 10 **by** 1" is generally written as "**for** I=1 **to** 10".

3. You must be careful while using the **for** statement lest you get in an infinite loop. For example, if you use

```
for I = 1 to 100 by -2
```

you will get into the infinite loop 1, −1, −3,....... Since the value of I is never going to exceed 100, the process would never terminate. On the other hand,

```
for I = 1 to 20 by 9
```

would result in action being taken for I=I, 10 and 19.

4. Generally, I is a subscript when used in a **for** statement. You are used to writing subscripted variables as $a_1, a_2, \cdots, a_n$ etc. and say "a_i for $i=1, 2, \cdots, n$". Here, we use the notation A[I] rather than a_i etc.

Example 2. *Write a program in pseudo-code to calculate the HCF of two positive integers.*

Solution. We can begin with problem-analysis.

1. What terms in the problem need to be defined or paid attention to ? HCF.

2. What are the inputs ? Two positive integers M and N.

3. What is going to be the output ? A positive integer.

4. What is the relation of the output to the input ? It is the HCF of the given numbers.

5. Are any methods to find HCF of two given numbers known ? Several ! In fact Euclid's algorithm in this connection is *famous.* A more modern approach is to write the prime factorizations of m and n as

$$p_1^{m_1} p_2^{m_2} \ldots p_k^{m_k} \text{ and } p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}. \text{ HCF}=p_1^{r_1} p_2^{r_2} \ldots, p_k^{r_k}$$

where $r_j = \min(m_j, n_j)$ for each $j=1, 2, \ldots, k$.

6. Which method shall be adopt ? Old is gold ! Let us use Euclid's algorithm ; the word *algorithm* sounds sweet.

7. What variables shall we need ? ...Let us see what we are going to *do* ; that will spell the variables for us. We begin by dividing the greater number N by the smaller number M. (That means, the first step is to be identify the smaller of the two given numbers.) If the remainder is zero, the smaller number which is the divisor, is the HCF. Otherwise, if R is the remainder, then 0⩽R<M. We now begin the same procedure a fresh starting with R as the smaller number or divisor and M (the earlier divisor) as the dividend. This spells out two things :

```
1.   get M, N
2.   Comment M, N are the given numbers.
3.   If M = N then put "HCF = "M
             else if M < N then SMALL ←— M ; BIG ←— N
                           else SMALL ←— N ; BIG ←— M
4    REMA ←— remainder from  BIG ÷ SMALL
5.   while REMA ≠ 0
           do BIG ←— SMALL
                     SMALL ←— REMA
                    REMA ←— remainder from BIG ÷ SMALL
6.   put "HCF = " SMALL
```

(*a*) *We are going to use repetition.* How long shall we iterate? Till the remainder is zero. Thus the condition to be tested for looping is "remainder=zero".

(*b*) Since every time, the divisor, the dividend, and the remainder get changed, three variacles are needed. Let us call the divisor SMALL, dividend BIG, and remainder REMA. When REMA is zero, the value of SMALL (the divisor) is the output HCF. The logic is complete now. The algorithm is given below.

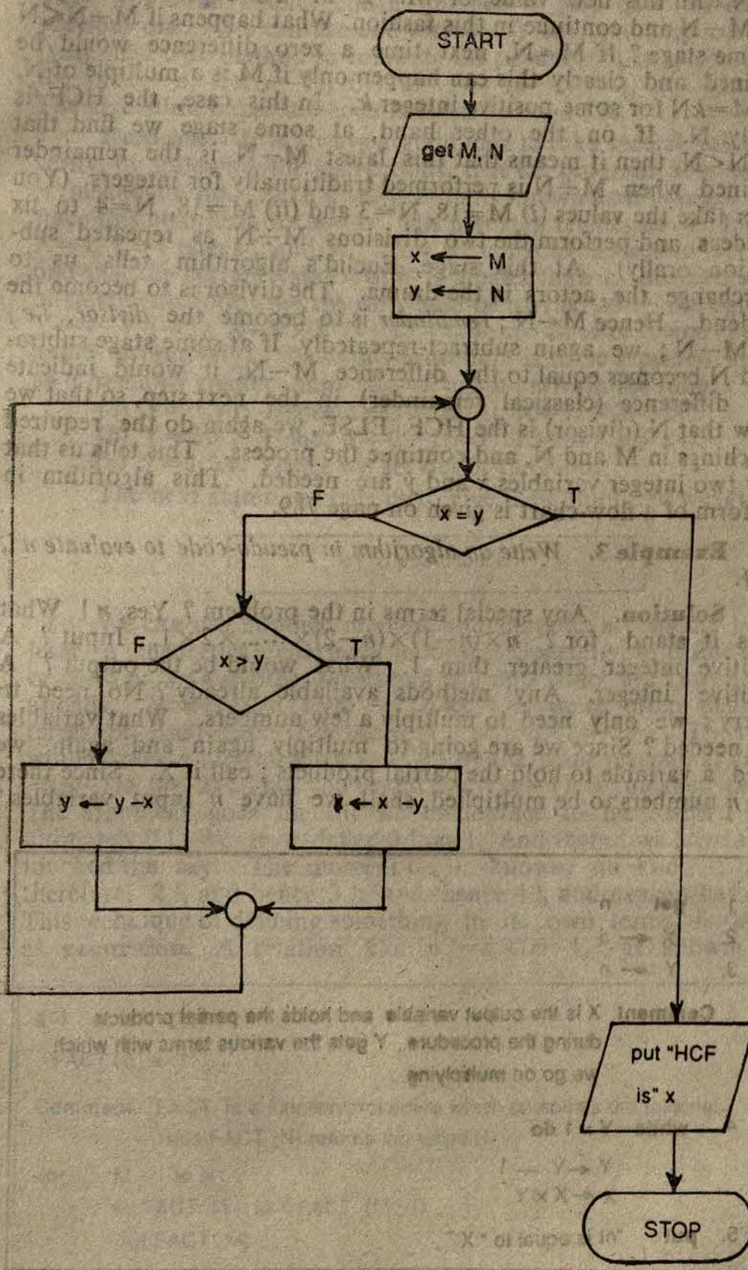
Remarks 1. Once we have written the algorithm, some imporovements generally become obvious. We could have cut down the first step easily. It won't hurt us even if we were dividing the smaller number by the bigger; it only gives a zero quotient and the smaller number as the remainder. Thus the traditional steps would start from the second step of our algorithm, but it would save us one nested loop. We could begin as follows :

get	M, N
BIG ←	M
SMALL ←	N
Comment	We do not worry if M is actually the smaller number; for all we care, M and N could be equal even.
•	• •
•	• •

The steps from fourth onward remain the same.

2. Generally, no algorithm is ready to be coded into a computer language, the first time it is written. Some steps always need *refinement*. For example, the assignment statement concerning REMA needs modification. How does a computer *divide*? Strictly speaking, it can only add and subtract. It carries out a division by repeated subtraction of the divisor from the dividend. Hence to divide M by N, it will first get M−N, then (M−N)−N, then ((M−N)−N)−N and so on, so forth, till the difference falls short of N. If we take this aspect into consideration, our algorithm would keep on doing M−N, suitably modifying the values of M and N all the time. This observation puts a new light on our method. First of all notice that so long as the task of 'dividing M by N' is not complete in the traditional sense, the various (M−N)'s above would be greater than N. In fact, the first M−N is O.K. Second time, we evaluate (M−N)−N. Thus we have made the assignment M←M−N; M has been modified. Again we perform

Flow-chart to compute the HCF of M and N



M−N with this new value of M. If M−N>N still, we again M←M−N and contiuue in this fashion. What happens if M−N⩽N at some stage ? If M=N, next time a zero difference would be obtained and clearly this ean happen only if M is a multiple of N, say M=kN for some positive integer k. In this case, the HCF is clearly N. If on the other hand, at some stage we find that M−N<N, then it means that this latest M−N is the remainder obtained when M÷N is performed traditionally for integers. (You better take the values (*i*) M=18, N=3 and (*ii*) M=18, N=4 to fix the ideas, and perform the two divisions M÷N as repeated subtraction orally). At this stage, Euclid's algorithm tells us to interchange the actors in the drama. The divisor is to become the dividend. Hence M←N ; *remainder* is to become the *divisor*, *i.e.*, N←M−N ; we again subtract-repeatedly. If at some stage subtrahend N becomes equal to the difference M−N, it would indicate zero difference (classical remainder) in the next step, so that we know that N (divisor) is the HCF. ELSE, we again do the required switchings in M and N, and continue the process. This tells us that only two integer variables x and y are needed. This algorithm in the form of a flow-chart is given on page 789.

Example 3. *Write an algorithm in pseudo-code to evaluate n !, n<1.*

Solution. Any special terms in the problem ? Yes, n ! What does it stand for ? $n\times(n-1)\times(n-2)\times\ldots\ldots\times2\times1$. Input ? A positive integer greater than 1. What would be the output ? A positive integer. Any methods available already ? No need to worry ; we only need to multiply a few numbers. What variables are needed ? Since we are going to multiply again and again, we need a variable to hold the partial products ; call it X. Since there are n numbers to be multiplied, shall we have n input variables ?

```
1   get    n
2.  X <-- n
3.  Y <-- n

Comment  X is the output variable and holds the partial products
         during the procedure.  Y gets the various terms with which
         we go on multiplying.

4.  while  Y > 1 do
           Y <-Y - 1
           X <-X x Y
5.  put    "n! is equal to " X ".
```

That would be stupid. Each number is just one less than the previous one. We can use one variable, Y say, and keep on modifying it. Our method to compute the product ? Let us compute in the order—n, $n\times(n-1)$, $(n\times(n-1))\times(n-2)$,......, $((\ldots\ldots)\times 2)\times 1$. The algorithm in pseudo-code is given on page 790.

Remark. It is always a good idea to take special values and verify, by following the instructions, that the program does what it is meant to do. Such a verification is known as **tracing**. Let us trace this algorithm for $n=4$. The first step fetches us 4. Initially, X=4, Y=4 (steps 2 and 3). Now we come to the **while** loop. We find that the current value 4 of Y is greater than 1. So the sequence of statements following the **do** would be carried out. This gives us Y←4−1 (=3), X←4×3 (=12) (current values of X and Y were 4 and 3 respectively). Now this procedure is to be repeated so long as Y remains greater than. So we again perform the loop.

(*a*) Y=3>1. Hence Y←3−1(=2), X←12×2(=24).

(*b*) Y=2>1. Hence Y←2−1(=1), X←24×1(=24).

(*c*) Y=1 ≯ 1. Hence we get out of the loop.

The next statement tells us to give the output and we get

```
4 ! is equal to 24
```

We can use another method to compute n ! which has, at least, a theoretical charm of its own. Generally, to define a term or concept, we use other terms and concepts which are known to us already. However, a term or concept is quite often defined in terms of itself too. For example, $n\,!=n\times(n-1)\,!$. Thus you would know n ! if you knew $(n-1)$! By the same reasoning, $(n-1)! =(n-1)\times(n-2)$! Thus you would know $(n-1)!$ if you knew $(n-2)$! The argument goes on till we come face to face with 1 ! and ultimately 0 !. We now define 0 ! as 1. And there ! we have almost touched the sky. The moment 0 ! is known, we know 1 !, and therefore, 2 !, and hence 3 !, and hence 4 !, and hence what not ! This technique of defining something in its own terms, is known as **recursion**, A relation like $n\,!=n\times(n-1)$! is known as a

```
get      N
  FACT (0) <- 1
Comment  FACT is a function/procedure which computes the factorial.
         Thus FACT (N) returns the value N!
for      M = 1 to N
         do FACT M = M x FACT (M - 1)
         put FACT (M)
```

recurrence relation. Recursive algorithms are easy to write but require a lot of storage and computer time at the time of implementation. There exist programs which convert recursive algorithms into non-recursive ones. Let us write a recursive algorithm to compute n !.

Example 4. (Fibonacci sequence) *In the beginning, there is one pair of rabbits, a male and a female. Two months later, the female rabbit starts giving birth to a like pair of rabbits every month. Each female rabbit behaves in this fashion. Write an algorithm to show the number of rabbit pairs at the end of each of the first K months. Assume no rabbits die.*

Solution. Clearly, at the end of the first two months there is only one rabbit pair. In the third month, the female gives birth and thus there are two pairs. During the fourth month, the first female produces another pair and there are three pairs at the end. During the fifth month, the second female also gets into the process of re-generation and two pairs are born this month. At the end of the fifth month, there are five pairs. Thus the first five terms in this sequence, known as Fibonacci's sequence, are 1, 1, 2, 3, 5. How many pairs would be there at the end of the nth month ? Let us call this number $T(n)$. Thus $T(1)=T(2)=1$, $T(3)=2$, $T(4)=3$ etc. There were $T(n-1)$ pairs at the end of the $(n-1)$th month and $T(n-2)$ at the end of the $(n-2)$th month. Now $T(n)=T(n-1)+$ rabbit pairs born during the nth month. How many rabbit pairs are born during the nth month ? Exactly as many as there were females of age at least two months. But this is equal to the number of rabbit pairs present at the end of the $(n-2)$th month. Hence

$$T(n)=T(n-1)+T(n-2). \qquad \ldots(A)$$

This gives us the required rule to generate the successive terms of the sequence or the required number of rabbit pairs.

Since we know the number of months in advance, we shall use the **for** statement. Using the above notation, T(1) and T(2) are to be assigned the value 1 each separately. For $n \geqslant 3$, we can use the relation (A) above to compute $T(n)$. The algorithm to compute and list the various terms is given below.

```
Get     K
T(1) ←  1
T(2) ←  2
for     N = 3 to K
        do
        T (n) = T (n-1) + T (n - 2)
put     T (n)
```

Example 5. *Write an algorithm for the solution of Brahma's Tower with N discs.*

Solution. As before, let us label the bars as ORIGIN, OTHER, and DESTINATION, but the discs as D[1], D[2],......, D[N]. As discussed earlier the following recursive method of solution may be adopted.

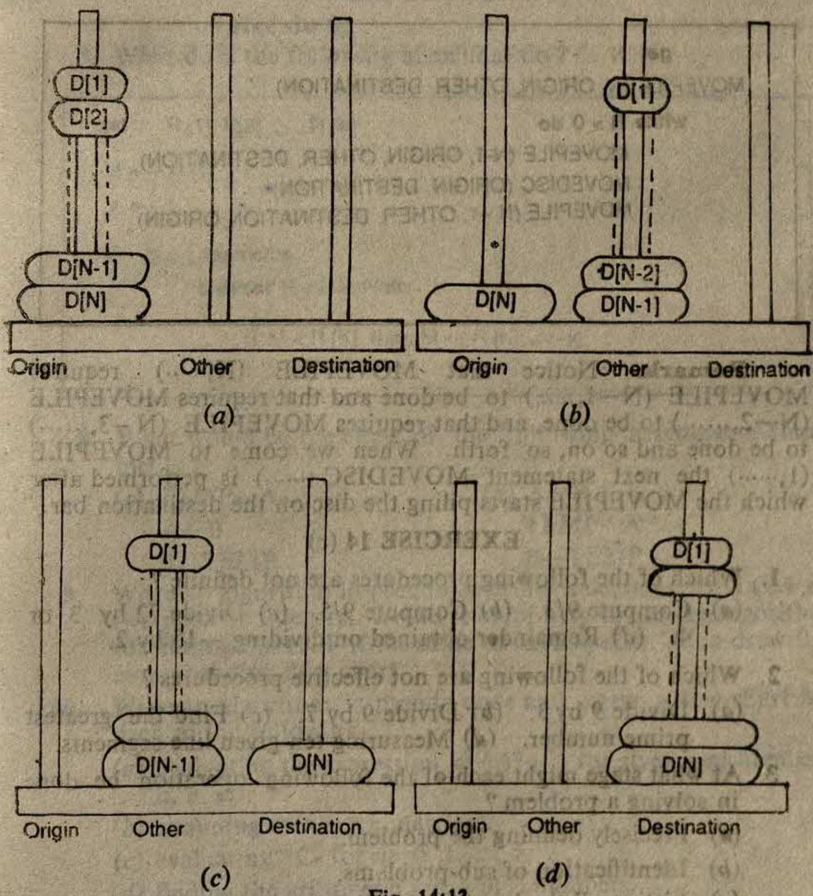


Fig. 14·13.

1. Move the top (N−1) discs from ORIGIN to OTHER, using the DESTINATION bar for intermediate piling, (Fig. 14·13(*b*)) Let us write this action as MOVEPILE (N−1, ORIGIN, OTHER, DESTINATION).

2. Transfer D[N] from ORIGIN to DESTINATION. (Fig. 14·13(*c*)). Let us write it as MOVEDISC (ORIGIN, DESTINATION).

3. Move the N−1 discs from OTHER to DESTINATION (Fig. 14·13(*d*)), using ORIGIN for intermediate piling, *i.e.*, perform MOVEPILE (N−1, OTHER, DESTINATION, ORIGIN).

The procedure consisting of the above three steps is labelled MOVEPILE (N, ORIGIN, OTHER, DESTINATION). The recursive program is written below :

```
    get N
MOVEPILE (N, ORIGIN, OTHER, DESTINATION)
    while N > 0 do
        MOVEPILE (N-1, ORIGIN, OTHER, DESTINATION)
        MOVEDISC (ORIGIN, DESTINATION)
        MOVEPILE (N -1, OTHER, DESTINATION, ORIGIN)
```

Remark. Notice that MOVEPILE (N,......) requires MOVEPILE (N−1,......) to be done and that requires MOVEPILE (N—2,......) to be done, and that requires MOVEPILE (N−3,......) to be done and so on, so forth. When we come to MOVEPILE (1,......) the next statement MOVEDISC (......) is performed after which the MOVEPILE starts piling the disc on the destination bar.

EXERCISE 14 (*c*)

1. Which of the following procedures are not definite ?
   (*a*) Compute 9/3. (*b*) Compute 9/5. (*c*) Divide 72 by 3 or 4. (*d*) Remainder obtained on dividing −15 by 2.
2. Which of the following are not effective procedures ?
   (*a*) Divide 9 by 3. (*b*) Divide 9 by 7. (*c*) Find the greatest prime number. (*d*) Measuring ten given line-segments.
3. At what stage might each of the following operation be done in solving a problem ?
   (*a*) Precisely defining the problem.
   (*b*) Identification of sub-problems.
   (*c*) Listing all the inputs.
   (*d*) Listing the desired outputs.
   (*e*) Devising suitable names for the occurring variables.
   (*f*) Exploring related algorithms.
4. Write the problem space for the algorithm of finding out loss or profit when SP and CP are given.
5. What is meant by structured programming ?

6. Draw a structured flow-chart for the following algorithm :

(*a*) **if** B **then do** S_1 ; S_2 ; S_3
else if C **then do** S_4 ; S_5
else do S_6

(*b*) **if** A **then if** B **then do** S_1 ; S_2
else while C **do** S_3
else do S_4

7. What does the following algorithm do ?

```
get   R [1], R[2], ...., R [N]
M ← R [1]
J ← 1
if  N = 1 then stop
       else for K = 2 to N do
            if M < R [K]  then M ← R [K]; J ← K
```

8. What would be the effect of the following assignment statements ?

(*a*) TEMP←A	(*b*) A←B
A←B	TEMP←A
B←TEMP	B←TEMP

9. Write an algorithm in pseudo-code to calculate the LCM of two positive integers. Would it be of help to use the algorithm for finding the HCF if the same is available ? Also draw the corresponding flow-chart.

10. Write an algorithm in pseudo-code and make a flow-chart for each of the following :

(*a*) evaluating the expression $a^3+b^3+c^3$ for given real numbers a, b, c.

(*b*) summing two $m \times n$ matrices.

(*c*) evaluating ${}^nC_r$ for $n=0, 1, ..., 5$.

(*d*) finding the arithmetic mean of n numbers.

SUMMARY

Computer (System)

A system consisting of a CPU and input-output devices. Or an electronic device which is capable of accepting data, processing these data automatically according to the stored instructions, and then giving out this processed data in the form of information.

Instruction

A statement which tells a computer to take some action.

Input

When encering Data fed into a computer for processing. (Also used as a verb.)

Arithmetic logic unit (ALU)

Component of CPU which performs the arithmetic and logic operations.

Control unit (CU)

The component of CPU which controls the sequence of actions in the computer.

Central processing unit

Component of computer consisting of ALU, CU and main storage.

Binary

A two-state system (represented by 0 and 1).

Bit

Abbreviation for a *binary digit* (0 or 1).

Code

A set of rules which transforms data from one representation into another.

Algorithm

An effective procedure which consists of a set of definite rules (steps), specifies a sequence of actions that provides the solution (output) to a given class of problems on the basis of one or more inputs in a finite number of steps.

Flow-chart

A graphical representation of an algorithm or the logic and the structure of a program.

Computer program

A set of instructions which tells the computer what to do.

Programming language

A set of words and their grammer, used to write computer programs.

Machine language

A low-level computer-dependent programming language made u p of zeros and ones intelligible to computer.

High level language

A computer language made up of words from natural languages, mathematical symbols and terminology of the problems it is meant to solve.

Simulation

Mimicking of the behaviour of one system by another ; technique to imitate part or all the behaviour of a system on the VDU.

Visual display unit

A TV-like device (equipped with a keyboard).

Ways to represent algorithms

(*a*) natural languages,
(*b*) decision tables,
(*c*) flow-chart,
(*d*) computer languages,
(*e*) pseudo-codes.

Structured programming

Style of writing algorithms in terms of the basic operations of sequence, selection and repetition.

Pseudo-code

A language having limited vocabulary and resembling both natural languages and computer languages.

REVIEW EXERCISE XIV

1. Describe the basic structure of a computer system.
2. Describe the functions of ALU.
3. Classify computer storage.
4. Write a short note on input-output devices.
5. Distinguish a typewriter from a KBD.
6. Explain the differences between the low and higher level computer languages.
7. What is ADA ? Who was ADA ?
8. Explain all the flow-chart symbols.
9. Draw the flow-charts corresponding to the three basic operations of sequencing, selection and repetition.
10. Give an example to show how the use of **case** statement may avoid nesting a loops.
11. What are the basic characteristics of a computer algorithm ?
12. What is meant by pseudo-code ?
13. **Odd ball.** In a collection of 9 similar looking balls, one ball is a little lighter. Show that two weighings are sufficient to identify the lighter ball. Write a suitable algorithm to do that. [**Hint** : Begin by putting 3 balls in each pan of the weighing scale.]
14. In the above problem, if it is not known as to whether the *odd* ball is a little lighter to heavier, then show that three weighings would suffice. Modify your algorithm to consider this situation.
15. Write an algorithm and draw the flow-chart for each of the following tasks :
    (*a*) listing the sixth row of Pascal's triangle.
    (*b*) matrix multiplication.
    (*c*) testing whether a given positive integer X is a prime.
    (*d*) generating 100 terms of a GP with common ratio R and first term A.
    (*e*) summing the squares of the first *n* natural numbers.
    (*f*) updating a bank account when a cheque has been drawn on the account.
    (*g*) arranging N given numbers in ascending order.

TEST YOUR UNDERSTANDING XIV

1. $(1111)_2+(1111)_2$ equals
   (*a*) $(11111)_2$, (*b*) $(1111)_2$,
   (*c*) $(10000)_2$, (*d*) $(11110)_2$,
   where $(\quad)_2$ indicates binary representation.
2. In a flow-chart, the statement "Add overtime to pay" would be enclosed in a
   (*a*) parallelogram, (*b*) rectangle,
   (*c*) diamond, (*d*) circle.

3. The statement "*variable expression*" is
   (*a*) a control statement
   (*b*) an input statement,
   (*c*) an assignment statement,
   (*d*) processing statement.
4. Initially, N=2 and A=3. The value of N after the execution of the three statements $N \leftarrow N+1$, $A \leftarrow A+N$, and $N \leftarrow N+N$, is
   (*a*) 5, (*b*) 9, (*c*) 6, (*d*) none of these.
5. A *selection* is made at
   (*a*) a function vertex, (*b*) a Boolean vertex,
   (*c*) a collecting vertex, (*d*) none of these.
6. A "**while**...**do**..." loop in a flow-chart represents the operation of
   (*a*) *sequence*, (*b*) *selection*,
   (*c*) *repetition*, (*d*) none of these.
7. A single loop in a flow-chart amounts to a **case** statement with
   (*a*) 0 alternative, (*b*) 2 alternatives,
   (*c*) 3 alternatives, (*d*) none of these.
8. For a computer algorithm to be *good* (efficient), which of the following should be maximum as against minimum ?
   (*a*) execution time on computer,
   (*b*) storage required on computer,
   (*c*) present and future flexibility for modification,
   (*d*) development time.
9. Of 25 similar looking gold coins, one is a bit lighter. The minimum number of weighings required to identify it is
   (*a*) 1, (*b*) 2, (*c*) 3, (*d*) 4.
10. The algorithm

```
EXPO ← 0
NUMBER ← 1
while      EXPO < 10  do
           NUMBER ← NUMBER * 2
           EXPO ← EXPO + 1
           print  NUMBER
Comment. * above indicates multiplication.
```

outputs

(*a*) 1, 2, ⋯, 10, (*b*) 2, 3, ⋯, 11,
(*c*) 2, 4, 6, ⋯, 20, (*d*) $2, 2^2, 2^3, \ldots, 2^{10}$.

HISTORICAL NOTE

Writing the *history* of something invented in *this* century might offend the Historians. Hence, computers are excluded from this note except to record their geneology.

First generation computers (early fifties) : These computers used *valve* for the two states 0 and 1. They were bulky and produced a large amount of heat. They were costly and failure-prone. They used punched cards for storage.

Second generatson computers (mid fifties to mid sixties) : Valves were replaced by transistors which were small, cheap and did not heat up quickly. Started using magnetic media. Became smaller, cheaper and more reliable.

Third generation computers (mid sixties to mid seventies) : Based upon integrated circuits (chips) instead of transistors. They *chipped in* a big way. Became much faster and cheaper. Software, however, becomes more sophisticated and more costly.

Fourth generation computers (mid seventies to early eighties) :

Based on very large scale integration (of chips). Hardware becomes yet cheaper. Software keeps the upward trend.

Fifth generation computers (under way) : Use ultra large integration. These are supposed to be *thinking* machines and have access to stored *expert knowledge*. Use *parallel* or *con current* mode of processing as against the usual *sequential* processing (one instruction at a time). What changes may they bring about ? Keep your fingers crossed ; wait and watch.

The word *algorithm* derives from a certain person named "Mohammed, father of Jafar, son of Moses and inhabitant of Khowarizm" (Abu Jafar Mohammed ibn Musa al-Khowarizmi). Khowarizm is known today as Khiva, and is a small city in USSR. Al-Khowarizmi wrote a book *Kitab al jabr wal-muqabala* (A book of rules of restoration and reduction). This book contained some rules regarding the arithmetic which Al-Khowarizmi had picked up from the Hindu works Al-Khowarizmi's name got associated with these *rules* and in time deformed into *algorithm*. Before the Greek idea of formal proof, most mathematics was *algorithmic*. Rules and procedures were stated to achieve certain goals without proof. Interest lay in getting the jobs done. Computers have turned the balance again. We are again getting *algorithmic* in style ; only there is a lot of hair-splitting as to how much computer time, storage, program code and so on would it take. Whether it is suitable for computer or not ; how to improve it, and so on, so forth. It is a really challenging and rewarding field to work in.

□ □

KARL FRIEDRICH GAUSS (1777-1855)

If there is a mathematician who can be regarded as a milestone on the long circuitous criss-cross road of mathematical evolution and revolution, it is Gauss, born in Brunswick, Germany on April 30, 1777. Mathematics before and after Gauss were much different because of the rigour introduced by him (and Cauchy and Abel). He is believed to be last man who knew all the mathematics of his time.

Gauss's entry into mathematics was almost providential. Had he not been able to construct a 17-gon with an unmarked ruler and collapsible (usable only once) compasses, he would have been a linguist. As it happens, the 17-gon decided Gauss in favour of mathematics. He put down this result in a small diary which ultimately became the most fruitful shortest diary of all times.

The roots of almost all modern numerical techniques are present in the works of Gauss in one form or the other. He was the first to deal systematically with the subject of numerical stability. His method of elimination of solving equations is yet the most popular.

As the years passed, Gauss surpassed everyone and everything in mathematics. His diary grew. In all, he filled 19 pages with 146 brief statements. Nevertheless these statements went to make a whole new world of mathematics. He published very little, for his motto was few but ripe. He created mathematics to satisfy his inner urge and let people have the credit for what he had done years earlier.

The king of Hanover justly ordered a medal which hailed Gauss as the *Prince of Mathematicians.*

Numerical Methods

15·1. INTRODUCTION

Do you like detective stories ? If *yes,* do you know *why* ? They say that people like detective stories becuase they are full of surprises. If that be true, you should like mathematics too, since it is so full of surprises ! Do you know that your usual laws of commutativity and associativity etc. in the context of addition and multiplication of real numbers may not be valid while you are working with your most sophisticated technological tools—the calculators and computers ? Would it come as a surprise to you that a computer cannot solve a simple equation like $x^2—2=0$ to get the roots as $\pm\sqrt{2}$? Moreover, that what it calculates as a *root* to all purposes and contents, may not satisfy the given equation, in the classical sense at least ? This chapter is aimed at presenting and solving mysteries like the ones mentioned here.

Recall that numbers like $\sqrt{2}$ which are a plaything for the mathematicians are nothing less than monsters to a computer. The reason is that all numbers must be represented as a finite string of symbols in a computer. Hence a number like $\sqrt{2}$ can never be represented *exactly* in a computer. We can only have as good an approximation to it as is feasible or desirable. This is one good reason why computer arithmetic can never be as exact as we are used to having so far. Yet, since there are important problems which can be solved only by means of computers, we relax the bounds of our accuracy a little and accept the *approximate numerical solutions* provided by computers as solutions. This brings in the questions like *how much error has crept into our solution* ; *how best to solve the problem in order to bring down the error to within certain limits*, and so on. Discussion of such considerations is what we call *numerical methods.* The key-words in this context are *approximations, error, numerical solutions, and algorithms which provide approximate solutions.*

15·2. DATA TYPES

You are aware that a computer is not merely a number crunching machine. It processes more non-numeric than numeric data. From the point of view of computers, all data are classified into the following three categories :

(*a*) *Numeric data, e.g.*, 37, $-24{\cdot}3$, $59{\cdot}7$, $3{\cdot}47\times10^5$ etc.

(*b*) *Alphabetic data, e.g.*, MEERUT, ANUPAMA, HAM SANDWICH THEOREM* etc.

(*c*) *Alphanumeric data, e.g.*, A–1 (a house number), DLJ 5716 (a car registration number), M 1101 (a bank account number) etc. These are a mixture of letters and numbers.

Recall that computer is essentlally a two-state machine. For that reason all types of data are stored in a computer by means of a string of two symbols, 0 and 1 say, each symbol representing one of the two states. Hence *all* types of data must be coded into strings of 0's and 1's. In this chapter, we shall be concerned with numeric data only. We shall first look into how such data are further classified, how they are represented in a computer, and how these representations may cause errors.

15·2·1. Numeric Data Types

There are two subdivisions of numeric data; *integer* and *real*. Integer data consist of the integers and may be *signed* or *unsigned*. For example, 3 may be written as +3 also. Written as '3', it is unsigned. Written as '+3' it is signed. Negative integers are always signed. Likewise, real data include real numbers, and may be signed or unsigned. These are expressed with the help of the decimal point.

15·3. REPRESENTATION OF NUMERIC DATA

As you know, the decimal system of notation is based on the concept of *place-value*. The magnitude of a digit is determined by virtue of the place it occupies with reference to the decimal point. The same combination of digits, say 1234, has different magnitudes according to the place it occupies with reference to the decimal point. For example, to the left of the decimal point as in 1234·0, it has a magnitude $1\times10^3+2\times10^2+3\times10+4\times10^0$; to the right as in ·1234, its magnitude is $1\times10^{-1}+2\times10^{-2}+3\times10^{-3}+4\times10^{-4}$; spread around as in 123·4, its magnitude is $1\times10^2+2\times10^1+3\times10^0+4\times10^{-1}$, and so on. In addition to assigning a particular value to a combination of digits as above, the decimal point also serves as thepole-star while we are carrying out various arithmetic operations. For example, consider the problem of adding 3·71 and 168·9. There is only one way to add these numbers. We *align the decimal point* as in Fig. 15.1 (*a*) before addition, and do not consider any other configuration like Fig. 15.1 (*b*) or 1 (*c*).

**Any three solids of any shape, placed anywhere in space, can be exactly halved by a plane* (like the slices of bread and one of ham in-between, when you cut it across to make sandwiches).

(a)	(b)	(c)
3·71	371	371
+168·9	+1689	+1689

Fig. 15.1

Again, consider the following addition problems :

71	7·1	·71
+26	+2·6	+·26
97	9·7	·97

Note that except for placing the decimal point at a particular place in the sum, the three problems are identical in that we only have to add 71 and 26 in all three cases. Thus so long as each number is properly aligned with the other number with respect to the decimal point, the configuration of the sum (97) remains the same. Hence we may treat all these three problems as one while performing addition, and worry about the decimal point later for each.

To make use of such economies as above, all numbers in computer arithmetic are given the same number of digits, the extra places being filled with zeros. For example, suppose we decide on a size of 6 digits for each number, four to the left and 2 to the right of the decimal point. Then 37 would be represented as 0037·00 ; 69·7 as 0069·70 ; ·89 as 0000·89, and so on.

Let us new see how various numbers are stored internally in a computer system. As you know, the main memory or the primary storage of a computer system generally consists of small ferrite cores which represent the binary 0 or 1 according to the direction (clockwise or counter-clockwise) in which they are magnitized. The ferrite cores are connected gridwise. A set of eight consecutive cores constitutes a *byte*. Thus a byte can represent eight bits. Depending upon the architecture, a computer *word* consists of 1, 2, 3, 4 or more bytes. A *memory location* is a collection of cores needed to represent one word Thus if a word is one byte (or 8 bit) long, a memory location consists of 8 consecutive ferrite cores. Different computer systems have different word-sizes.

Memory locations are numbered 0 upwards. These numbers by which these locations are recognized, are known as their *addresses*. For the sake of convenience, we shall assume that a memory location of our computer system consists of eight bits or one byte. We shall use the following diagram consisting of 8 consecutive cells to represent one memory location even though the actual physical picture, *i.e.*, the hardware component called a memory location, is rather different.

One unit of data is stored in one memory location. The left most bit is left for the sign. We may use a *zero* for the *plus* sign and a *one* for the *minus* sign. This convention may be different for different systems. The integers are written *right justified, i.e.*, we start writing the bits in the $2^0, 2^1, 2^3, \ldots\ldots$ positions from the right-most cell, filling the unused cells on the left, if any, by zeros. Look at the following illustrations to fix the idea.

0	0	0	0	0	0	0	1	$\longleftrightarrow (1)_{10} = (1)_2$

Sign bit 0 denoting a positive number

1	0	0	0	1	1	1	0	$\longleftrightarrow (-14)_{10} = (-1110)_2$

Sign bit 1 denoting a negative number

How many digits have been assigned to each number in the above representation ? Seven ! Where is the point ? (Binary point in this case.) Not *shown*. It is *implied* at the right-most position after the 2^0 bit. The same convention as we adopt while writing the integers. However, the interesting point is that we may *assume* it at any position (without showing) and fix this position once for all throughout a particular working phase. Thus suppose we fix it to the left of the two right-most cells as shown below :

implied decimal point

Then the numbers stored above would be interpreted as

$$(+00000{\cdot}01)_2 \text{ or} \left(\frac{1}{4}\right)_{10},$$

and $(-00011{\cdot}10)_2$ or $\left(2^1+2^0+\dfrac{1}{2}\right)_{10}$.

Such a representation clearly imposes a servere limitation on the size of numbers we can use on a computer. The biggest number that we can store in our 8-bit word above (or a memory location) is $(1111111)_2$. The decimal equivalent of this binary number is

$$1\times2^6+1\times2^5+1\times2^4+1\times2^3+1\times2^2+1\times2^1+1\times2^0$$
$$=2^6+2^5+2^4+2^3+2^2+2+1=2^7-1=127.$$

Obviously, this is a deplorable situation. We certainly need to work with bigger numbers. Even if we were to increase the word length, not much would be attained. Thus for example, had our word been 2 bytes long, the largest decimal integer that we could store would have been $2^{15}-1$ (or 32767). The largest positive integer representable by a 4 bytes long word would be 21, 47, 483, 647, which is again inadequate even though the hardware complications have increased, not merely four-fold but several-fold because of the bigger word size. Since particular place, the above representation of real numbers is known as the **fixed point representation**.

15·3·1. Floating Point Representation of Real Numbers

As you must have realized by now, the fixed point representation of real numbers may not always be adequate or most convenient. At times, we may be talking about very big or very small numbers like the mean distance between the sun and the earth which is 14,950,000,000 km., or the radius of an electron which is 0·00000000000281784 mm. You are aware that one convenient way of writing such clumsy numbers is to use the scientific notation. Thus we could say that the mean distance mentioned above is $1·495\times10^7$ km. This expression suggests some other ways too of writing the above number. We could as well write this number as $·1495\times10^8$ or $14·95\times10^6$ or $·01495\times10^9$ or $149·5\times10^5$ etc. Notice how we can *float* the decimal *point* this way or that by making an adjustment in the exponent of 10. Such a representation of real numbers is accordingly known as the **floating point representation**. Clearly, all real numbers (and not necessarily too big or too small only) have (more than one) floating point representations. For example, three of the floating point representations of the fixed point real number 37·0 are $·37\times10^2$, $3·7\times10^1$ and $370·0\times10^{-1}$.

Notice that a floating point representation of any real number has the typical form (a fixed point real number)$\times10^n$, where n is an integer. In this representation, the fixed point real number part is called the **mantissa** or the **argument**, 10 is called the **radix** or **base**, and n is called the **exponent**. The base is generally understood and thus the number is defined by its mantissa and exponent.

Illustration. In the number $67{\cdot}5\times 10^{-3}$, the mantissa (or argument) is 67·5, the base (or radix) is 10, and the exponent is -3.

Remarks 1. The mantissa tells you *what* the number is and may be a negative real number.

2. The exponent is an integer and may be negative or zero also. It tells you *where* the decimal point is.

3. The base being 10, division and multiplication by powers of 10 can be effected merely by *shifting* the decimal point to and fro in the number. Equivalently, if we wish to keep the decimal point fixed, this amounts to shifting the digits right and left. For example, to multiply 3·79 by 10, we *shift the decimal point one place to the right*, getting 37·9. This is the same thing as keeping the decimal point fixed and *shifting all the digits one place to the left.* Similarly a multiplication by 10^n $(n>1)$ is effected by shifting the

Effect of multiplication by 10

3·79 (before) 37·9 (after)	3·79 (before) 37·9 (after)
Digits in place ; *decimal point* *shifted right.*	*Decimal point in place ;* *digits shifted left.*

decimal point n places to the right or by shifting all the digits n places to the left. For division by powers of 10, *left* and *right* in the above thumb rule are interchanged.

Raising the exponent means multiplying the number by a positive power of 10. Thus if we wish to modify the exponent of a real number in its floating point representation, but wish to keep the value of the number unchanged, we must divide its mantissa by the same power of 10. For example,

$$\cdot 1234\times 10^{5}=\cdot 01234\times 10^{6}, \text{ since } \cdot 1234\times 10^{5}=\cdot 1234\times\left(\frac{10^{6}}{10}\right),$$
$$=\frac{\cdot 1234}{10}\times 10^{6},$$
$$=\cdot 01234\times 10^{6}.$$

Similarly, lowering the exponent is countered by multiplying the mantissa by a suitable power of 10*. For example $\cdot 1234\times 10^{5}=123{\cdot}4\times 10^{2}$ since

$$\cdot 1234\times 10^{5}=\cdot 1234\times(10^{2}\times 10^{3})=(\cdot 1234\times 10^{3})\times 10^{2},$$
$$=123{\cdot}4\times 10^{2}.$$

*Raising (resp. lowering) the exponent by n (>0) means shifting the decimal point in the mantissa n places to the left (resp. right).

Since a real number can have several different floating point operations, it would be a good idea to agree upon a standard one. Since the decimal point can be floated at will for all non-zero numbers, we can always adjust it so that it is placed to the left of the left-most non-zero digit in the mantissa, thus making the mantissa a proper fraction, or a number whose magnitude lies between 0 and 1. Such a floating point representation is called the **normalized representation**. Zero cannot be written in the normalized form because it does not have a non-zero digit in its mantissa. Written in the normalized form, the mantissa gives the *significant digits* in the number and the exponent gives its *size* or *scale*.

Illustration. The numbers $\cdot 32791\times10^{-5}$, $\cdot 1411\times10^{3}$, $\cdot 3542\times10^{0}$, $-\cdot 67\times10^{4}$ are all in the normalized floating point form. None of $3\cdot 2791\times10^{6}$, $14\cdot 11\times10^{1}$, $354\cdot 2\times10^{-8}$ and $-6\cdot 7\times10^{3}$ is a normalized floating point representation.

Floating point representations are expressed in the following format :

Mantissa E Exponent

.For example, we write $\cdot 375$ E+14 for $\cdot 375\times10^{14}$, $\cdot 1492$ E-7 for 1492×10^{-7}, $-\cdot 199$ E+5 for $-\cdot 199\times10^{5}$ and $-\cdot 234$ E-2 for $-\cdot 234\times10^{-2}$. Notice that (*i*) the base (10) being understood, has been omitted in the above representation, (*ii*) the mantissa which is a fraction, may be negative, and (*iii*) the exponent, separated from the mantissa by the letter E, is a signed number.

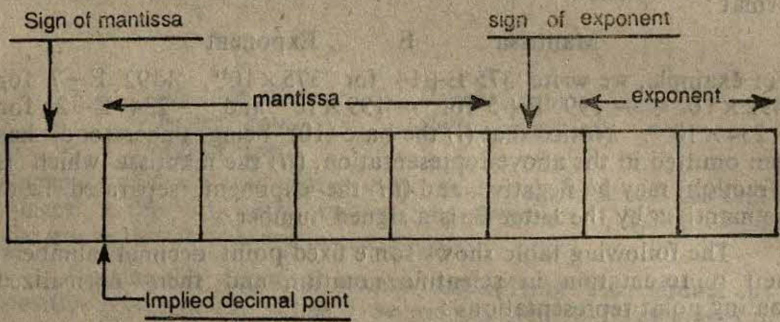
The following table shows some fixed point decimal numbers, their representation in scientific notation and their normalized floating point representations :

Fixed point decimal form	Scientific notation	Normalized floating point representation
0.003597	3.597×10^{-3}	.3597 E -2 (= $.3597\times10^{-2}$)
27.69	2.769×10^{1}	.2769 E 2 (= $.2769\times10^{2}$)
3421	3.421×10^{3}	.3421 E 4 (= $.3421\times10^{4}$)
-0.000034	-3.4×10^{-5}	$-.34$ E -4 (= $.34\times10^{-4}$)

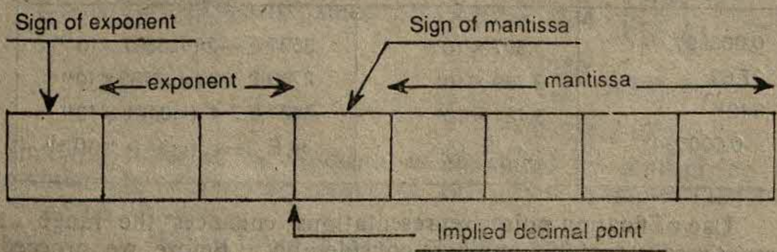
Use of floating point representations enhances the range of computer-representable numbers considerably. Before we proceed to see how and to what extent this can happen, let us appreciat one fact. All numbers handled by any computer system are ultie mately in a binary format. Yet, whether we use base ten or base two, the *procedures* as such do not change. It is only the *representations* of the numbers which are different. So far as the end-

users such as you and we are concerned, we may never have to enter our data in a binary code. All the output that we would receive would also be in the decimal code only. Hence now onwards we shall generally explain the concepts using base ten only. To begin with, let us learn how floating point representations are stored in the system. Only normalized floating point representations are used.

Recall that a normalized floating point representation (NFPR henceforth) consists of two parts : the fraction part or the mantissa, and the exponent. How numbers are represented in the computer system depends upon the word-size and varies from machine to machine. Let us assume that a word consists of a byte and as before, let us use the left-most bit for the sign of our number. Of the remaining seven, the next four may be used to write the mantissa ; the next in order for the sign of the exponent, and the remaining right-most two bits for the exponent as shown below. Just as the base is understood, so the position of the decimal



point is understood. More often, the following representation is used :



We shall, however, stick to the earlier representation which is more suggestive because of the way we express FPR of real numbers. Also, we shall use plus, minus signs instead of 0 and 1 respectively for them.

Illustrations.

Stored form								*NFPR*
+	3	4	7	2	+	1	0	$\cdot3472E+10\ (=\cdot3472\times10^{10})$
+	3	4	7	2	–	1	0	$\cdot3472E-10\ (=\cdot3472\times10^{-10})$
–	9	1	0	0	+	0	3	$-\cdot9100E+03\ (=-\cdot9100\times10^{3} = -\cdot91\times10^{3})$
–	7	2	1	0	–	0	1	$-\cdot7210E-01\ (=-\cdot721\times10^{-1})$

Let us now examine what is the range of the magnitudes of the numbers we can represent in the above fashion. The greatest and the least possible positive maintissas are clearly $\cdot9999$ and $\cdot1000$ because the most significant digit in the mantissa has to be non-zero. The range of the exponent, including the sign, is -99 to 99. Thus the smallest positive real number which can be represented as above is $\cdot1000\times10^{-99}$ and the biggest one is $\cdot9999\times10^{99}$. As you can see, these are astronomical numbers and we may need never go outside this range for reasonable practical applications. The negative numbers which can be represented on this system in this way are the numbers in the range $-\cdot9999\times10^{99}$ to $-\cdot1000\times10^{-99}$. What about the numbers in the range $-\cdot1000\times10^{-99}$ to $\cdot1000\times10^{-99}$? How unfortunate! these numbers cannot be represented here. Come to think of it, what are we losing anyway? A number in this range is going to have 99 zeros after the decimal point. Such a small number has to be treated as zero for all practical purposes.

There is yet another difficulty. Consider the number $345678\cdot9$. We can write it as $\cdot3456789\times10^{6}$ so that the NFPR of this number is $\cdot3456789\ E+06$. How shall we store it in our computer system? It certainly lies in the range calculated above. Since we have only four cells to represent the mantissa, not all the digits of the mantissa $\cdot3456789$ can be accommodated on our computer system. What it would do is to store the four most significant digits 3, 4, 5 and 6, and would *chop off* the remaining digits 7, 8 and 9. Thus this number would be stored as follows :

3456789E + 06 ⟷

+	3	4	5	6	+	0	6

The stored number $\cdot3456\times10^{6}$ is only an approximation of the given number $\cdot3456789\times10^{6}$ and thus a certain error has crept into our data. To be precise,

$$\begin{aligned}\text{Error} &= \text{Actual number} - \text{stored number},\\ &= {}^{\cdot}3456789 \times 10^6 - {}^{\cdot}3456 \times 10^6,\\ &= 10^6({}^{\cdot}3456789 - {}^{\cdot}3456),\\ &= 10^6 \times {}^{\cdot}0000789 = 78{\cdot}9.\end{aligned}$$

Compared to the number, this is small because

$$\frac{\text{Error}}{\text{Actual number}} = \frac{10^6 \times {}^{\cdot}0000789}{10^6 \times {}^{\cdot}3456789}$$

$$= \frac{789}{3456789} (= 0002 \text{ or } {}^{\cdot}02\% \text{ approx.})$$

For all practical purposes, an error of 789 in a number of the magnitude of 3456789 can be ignored. Nevertheless, the error is there ; and don't you be under the impression that a bigger word-size would eliminate the error completely. It can only *reduce the error or increase the precision*. Numbers like $\sqrt{2}$ which happen to be non-terminating decimals cannot be expressed exactly by a word of *any* length. So, even the most powerful of computer systems have their limitations. We shall have more to say about errors in the next section.

EXERCISE 15 (*a*)

1. Write down the mantissas of each of the real numbers whose floating point representations are given below :
   (*a*) 3·7961 E+02. (*b*) 2·00095 E+04.
   (*c*) ·0079 E−02. (*d*) −·151 E+01.
   (*e*) −33·149 E−03. (*f*) −5496 E−04.
2. Write down the exponents of each of the numbers given in problem 1 above.
3. Express each of the numbers given in problem 1 in the fixed point form.
4. Exress each of the following in a floating point form with exponent 4 :
   (*a*) 3742·6. (*b*) 3742·65.
   (*c*) 12345. (*d*) 1·2345.
   (*e*) −12·345. (*f*) 1234·5.
5. Express each of the numbers in problem 4 above with exponent −3.
6. For each of the following, give that floating point representation in which the mantissa is equal to 12·34 :
   (*a*) ·01234. (*b*) ·1234.
   (*c*) 1·234. (*d*) 12·34.
   (*e*) 123·4. (*f*) 1234.
   (*g*) 12340. (*h*) 123400·0.

7. Express in the normalized floating point form :

(*a*) 372 E 5. (*b*) ·372 E 5.
(*c*) 37·65 E −5. (*d*) 9·99 E 9.
(*e*) 1234. (*f*) ·0057.
(*g*) 6·1. (*h*) ·2345. (*i*) ·0001.

8. How would the following numbers be stored in the computer mentioned in the text ?

(*a*) ·2345 E+12. (*b*) 2·345 E 12.
(*c*) 23·45 E−12. (*d*) 259·11 E+09.
(*e*) −359·79 E−2. (*f*) 29·2 E 8.

Which of these when stored are in error ? Calculate the actual errors. Also express the errors as compared to the given numbers in the form of a per cent.

15·5. BASIC OPERATIONS WITH FLOATING POINT REPRESENTATIONS

You have seen that in order to be able to work on computers with numbers in a reasonable range, we must use their floating point representations. We shall now try to construct some algorithms to find the sum, difference, product and quotient of such numbers without feeling compelled to first converted them into the fixed point form.

15·4·1. Addition of Real Numbers in NFPR

Recall how you add two decimal numbers like 13·79 and 2·794. You first align the decimal point, then add ignoring the decimal point, and finally put the decimal point in the sum so as to align it with the decimal point of either summand. Recall that in floating point operations, *it is the exponent which tells you where the decimal point is*. Hence aligning the decimal point amounts to making the exponents of the summands equal by adjusting the mantissa of one of the numbers. Which one ? We shall soon answer that. Adding the summands now (without bothering for the decimal point) means adding the *mantissas which tell you what the number is*. Aligning the sum with the summands amounts to making their exponents equal. Let us fix these ideas by taking the following addition sums :

(*a*) $\cdot 2314\times 10^5+\cdot 3112\times 10^5$

(*b*) $\cdot 9710\times 10^2+\cdot 2311\times 10^2$

(*c*) $\cdot 3714\times 10^5+\cdot 215\times 10^3$

(*a*) The sum presents no difficulty on the machine. Already, the exponents are equal. We would have carried out the operation as follows :

$$\cdot 2314\times 10^5+\cdot 3112\times 10^5=(\cdot 2314+\cdot 3112)\times 10^5$$
$$=\cdot 5426\times 10^5.$$

The floating point representation would have been

$$\cdot 2314 \text{ E}+05+\cdot 3112 \text{ E}+05=\cdot 5426 \text{ E}+05.$$

The answer would have been stored as follows :

+	5	4	2	6	+	0	5

(*b*) Like (*a*), the exponents are equal here also. However, when the mantissas are added, we get 1·2021. Now the memory location which is going to hold the sum has only four bits for the mantissa and cannot hold such a big mantissa as 1·2021. Such a situation is described by saying that there is an *overflow* in the mantissa. An overflow arising in this manner during the working is taken care of by the circuitry of the machine and brings to the fore another advantage of the floating point representations. In cases like this the scaling is *automatically done*, *i.e.* the machine re-writes a number like 1·2021 E 2 automatically as ·1202 E 3. In other words, the exponent of the sum is raised by one and the mantissa shifted right to one place. In this process, the least significant digit is lost. For example,

$$1\cdot 2021 \text{ E } 2=\cdot 12021 \text{ E } 3, \text{ stored as } \cdot 1202 \text{ E } 3,$$

the least significant 1 being chopped off. Thus we have the following :

$$\begin{array}{r} \cdot 9710 \text{ E } 2 \\ +\cdot 2311 \text{ E } 2 \\ \hline 1\cdot 2021 \text{ E } 2=\cdot 1202 \text{ E } 3. \\ \hline \end{array}$$

The answer would be stored as below :

+	1	2	0	2	+	0	3

Remark. Quite likely, on having encountered an overflow in the mantissa, a machine would scale the numbers themselves before summing them, that is, would raise the exponents of both the summands by 1 by shifting the mantissas of both the summands one place right. This would result in the loss of the least significant digit of both the summands. For example, the following operations would be triggered as soon as an overflow is encountered :

$$\cdot 9710 \text{ E}+02 \leftrightarrow \cdot 0971 \text{ E}+03,$$
$$\cdot 2311 \text{ E}+02 \leftrightarrow \cdot 0231 \text{ E}+03,$$
$$(\cdot 0971 \text{ E}+03)+(\cdot 0231 \text{ E}+03) \leftrightarrow (\cdot 1202 \text{ E}+03)$$

The sum obtained here is the same as before. Construct an example where the two sums obtained in these two ways are NOT the same. Which way do you get a *less erroneous* answer? The *way* you do things is *very important* as we shall see in section 15·5.

(*c*) We could proceed in two different ways. We could either make the exponent of both the terms 3, or we could make the exponent of both the terms 5 by a suitable modification in the mantissa of one of the terms. If we make the exponent 3 common, the first term $\cdot 3714 \times 10^5$ must be written as $37 \cdot 14 \times 10^3$. The mantissa being stored as a fraction, 37 here would be lost. That is, the most significant digits would be lost. Hence this approach is no good. We should, therefore, raise the lower exponent rather than lower the higher exponent to make the exponents of both the summands equal. Thus, *keeping the term with higher exponent as it is and modifying the other*,

$$\begin{array}{r} \cdot 3714\text{ E}+05 \\ +\cdot 2150\text{ E}+03 \\ \hline \end{array} \quad \leftrightarrow \quad \begin{array}{r} \cdot 3714\text{ E}+05 \\ +\cdot 0021\text{ E}+05 \\ \hline \cdot 3735\text{ E}+05 \\ \hline \end{array}$$

The above analysis gives us the following thumb rule for adding two NFPR of real numbers :

Step 1. If the exponents of the summmands are different, then *raise the lower exponent to become equal to the higher one* by a suitable modification in the mantissa (which may involve the discarding of some least significant digits).

Step 2. To find the mantissa of the sum, add the mantissas of the two summands. If the sum is less than 1, then this is the required mantissa. Else divide this sum by 10, *i.e.*, shift it one place right, keeping only the four most significant digits. This is the required mantissa.

Step 3. Take the common exponent of the summands as the exponent of the sum if no adjustment in the mantissa of the sum was made. Else raise the common exponent by 1. This is the exponent of the sum.

Remark. Since we have reserved four digits for the mantissa, two for the exponent, and the computer words are of fixed length, it would be a good practice to write all our numbers accordingly. For example, $+\cdot 3700$ E$+06$ rather than $\cdot 37$ E 6. However, for brevity and convenience, we shall at times use the latter from. When numbers are stored in the computer, then of course things would be different.

Example 1. *Add ·2002 E 5 and ·7991 E 5.*

Solution. The exponents being equal, the first step above is skipped. The sum of the mantissas is ·9993. This is the mantissa of the sum. The exponent of the sum is 5. The whole working is arranged as follows :

$$\begin{array}{r} \cdot 2002\text{ E }5 \\ +\cdot 7991\text{ E }5 \\ \hline \cdot 9993\text{ E }5 \end{array}$$

Example 2. *Find the sum of ·5391 E−3 and ·7123 E−3.*

Solution. We notice that the exponents are identical, so step 1 is skipped. To get the mantissa of the sum, we add the mantissas of the two summands. This gives us 1·2514 which is greater than 1. To make it a fraction, we divide by 10 (or keeping the decimal point fixed, shift all the digits one place to the right). This give us ·12514. To keep only four digits in the mantissa, we discard the least significant digit 4, thus getting ·1251 as the mantissa of the sum. Since an adjustment has been made in the mantissa of the sum, we raise the common exponent of the summands by 1, thus getting the exponent of the sum as −2. The actual working is quite straight-forward and can be arranged as follows :

$$\begin{array}{r} \cdot 5391\text{ E }-3 \\ +\cdot 7123\text{ E }-3 \\ \hline 1\cdot 2514\text{ E }-3=\cdot 1251\text{ E }-2 \end{array}$$

Remarks 1. *Raising the exponent by* 1 means *adding one to it.* Thus $-3+1=-2$. You didn't want to make it −4 ! or did you ?

2. If the sum exceeds the biggest number which can be stored, then there would really be an overflow which cannot be adjusted by scaling. For example, the highest number which we can store in our assumed computer is ·9999 E+99. The sum of ·9995 E+99 and ·1234 E+99 exceeds this number. To indicate this *condition of overflow*, an error message of some sort would be obtained so that the programmer may do something about it. So long as the overflow in mantissa can be accommodated by scaling, no error message is given out.

Example 3. *Evalute (·1234 E+05)+(·9870 E+06).*

Solution. Here, the exponents are different. The lower exponent is 5. It is to be raised to 6. Hence we shall keep the number with exponent 6 as it is, but modify the other so as to make its exponent 6. Now

$$\cdot 1234\text{ E}+05=\cdot 0123\text{ E}+06,$$

discarding the least significant digit 4 in the original mantissa to keep four digits in the new mantissa. Thus

$$\begin{array}{r} \cdot 1234\text{ E }5 \\ +\cdot 9870\text{ E }6 \\ \hline \end{array} \leftrightarrow \begin{array}{r} \cdot 0123\text{ E }6 \\ +\cdot 9870\text{ E }6 \\ \hline \cdot 9993\text{ E }6 \end{array}$$ **Ans.**

EXERCISE 15 (*b*)

1. Add and store the answer in the normalized form assuming provision for four digits in the mantissa and two in the exponent :

   (*a*) ·3124 E 10 and ·6875 E 10.
   (*b*) ·5432 E 19 and ·3456 E 19.
   (*c*) ·6556 E 5 and ·1221 E 4.
   (*d*) ·8888 E 4 and ·1111 E 2.
   (*e*) ·2431 E −5 and ·1234 E −5.
   (*f*) ·6741 E −4 and ·2031 E −4.
   (*g*) ·1651 E −5 and ·1098 E −6.

   [Be careful about the higher exponent ; it is not −6.]

   (*h*) ·3773 E −8 and ·7337 E −6.
   (*i*) ·9876 E 1 and ·1165 E −1.
   (*j*) ·2365 E −1 and ·6708 E 2.
   (*k*) ·9996 E −1 and ·8808 E 3.

   [See ! what a funny object computers are !]

   (*l*) ·7342 E −5 and ·9801 E 1.
   (*m*) ·6402 E 7 and ·7892 E 7.
   (*n*) ·9001 E 5 and ·6903 E 5.
   (*o*) ·9804 E 94 and ·9872 E 93.
   (*p*) −·6402 E 7 and −·2203 E 7.

2. Suppose there were provision for eleven digits in the mantissa and three in the exponent. Which of the sums in problem 1 above would change ? Find the new sums in their NFPR.

3. Which of the following would cause an error message on account of an overflow condition in a machine with provision for two digits in the exponent and four in the mantissa ?

   (a) ·9999 E 98+·1101 E 98.
   (*b*) ·9999 E 98+·8889 E 98.
   (*c*) ·1013 E 99+·7824 E 99.
   (*d*) ·2913 E 99+·7824 E 99.
   (*e*) ·2913 E 99+·9999 E 98.
   (*f*) ·7654 E 99+·7009 E 99.
   (*g*) ·9999 E 99+·9999 E 95.

15·4 2. Subtraction with Normalized Floating Point Representations of Real Numbers

Would you believe that it is possible to subtract *any* number from *any* number without ever borrowing ? In particular can you subtract 57 from 63 without borrowing ? Believe it or not, but the answer is *yes*. Let us see how.

$$\begin{aligned} 63-57 &= 63+(100-57)-100, \\ &= 63+\{(99-57)+1\}-100, \\ &= 63+42+1-100, \text{ (no borrowing !)} \\ &= 106-100=6. \text{ (no borrowing !)} \end{aligned}$$

Clearly, the above technique would work for any two integers. The number (99—57) is known as the **9's complement** of 57. The 9's complement of a non-negative integer is obtained by replacing every digit in the integer by '9 minus that digit'. For example, 9's complement of 20 is 79 which is obtained on replacing 2 by $9-2$ and 0 by $9-0$ in 20.

The number $(99-57)+1$ is known as the **10's complement** of 57. The 10's complement of any non-negative integer is obtained by adding 1 to its 9's complement. Thus for example, the 10's complement of 61 is $(99-61)+1$ or 39. In effect, the ten's complement of an n-digit integer is nothing but 10^n minus that integer, but we *do not* want to calculate it that way. It is the problem of *subtraction* that we are trying to convert into a problem of addition. Subtraction is carried out in a computer by adding the 2's* complement of the subtrahend (the number to be subtracted) to the minuend (the number from which we have to subtract) and then dropping the most significant digit of the sum. This dropping is the equivalent of subtracting the number which was added while forming the 2's complement. For example, the computer would subtract 11010 from 11100 in the three steps shown below.

I. *Finding the 2's complement of the subtrahend* 11010. This is obtained by adding 1 to the 1's complement of this number. The 1's complement is obtained by replacing every bit in the number by '1 minus the bit'. Since $1-0=1$ and $1-1=0$, this amounts to replacing every zero by a 1 and every 1 by a zero. Thus 1's complement of 11010 is 00101. *No subtraction performed.* The algorithm giving the 1's complement is : *Interchange the zeros and the ones.* Hence the 2's complement of 11010 is 00110.

II. Adding the minuend and the 2's complement of subtrahend. This gives $11100+00110=100010$.

---

* The role of the decimal ten is played by the binary 10 (*i.e.*, decimal 2) in the binary and that of decimal nine (the biggest digit in base 10 system) by binary one (the biggest bit in binary system).

III. *Dropping the most significant bit in the sum.* This gives 00010 or 10 as the required answer.

Subtraction can be carried out by using the 1's complement also. In this case, we add the minuend and the 1's complement of the subtrahend. The 1 which was added in forming the 2's complement in the previous case, remains to be added to this sum. The 1 which occurred as the most significant bit remains to be dropped. So our most obedient servant, the computer is instructed to carry off this latter 1 (knokn as the **end-around carry**) and add it to the sum formed, thus producing the required difference.

Example 4. *Subtract* $(11011100)_2$ *from* $(110111100)_2$ *by* (*i*) *using the 1's complements.* (*ii*) *using the 2's complements.*

Solution. The subtrahend is 11011100. To get its 1's complement, replace every 1 by 0 and conversely. This gives 00100011. To get 2's complement of the subtrahend, add 1 to its 1's complement. This gives 00100100.

(*i*) 110111100	(minuend)
+000100011	(1's complement of subtrahend)
+111011111	(temporary sum)
+————→1	(removing the end-around carry and adding it into the sum thus deleted)
11100000	(final sum, *i.e.*, the required difference)
(*ii*) 110111100	(minuend)
+000100100	(2's complement of subtrahend)
(1)11100000	(temporary sum)
11100000	(final sum or the required difference)

Subtraction problems involving floating point representations are also done (in computers) by converting them into addition sums. Recall that adding 6·7 and 2·1 is virtually the same as adding 67 and 21 etc. when we are using floating point representations. We shall now see how subraction is carried out by using the 9's complements and 10's complements in the decimal system. Finding the complements was easy in the binary system. Here our algorithm would involve 10 subrules like '1's complement of 9 is 0', *i e.*, 'replace 9 by 0', 'replace 8 by 1', ..., 'replace 0 by 1' etc. We shall use the notation '$a \leftarrow b$' for 'a is replaced by b'.

Example 5. *Subtract 3764 from 5129 by using 10's complements.*

Solution. For finding the 9's complement, the following replacements are made in 3764 :

$$3 \leftarrow (9-3)\ ;\ 7 \leftarrow (9-7)\ ;\ 6 \leftarrow (9-6) : 4 \leftarrow (9-4).$$

Thus 9's complement of 3764 is 6235. Hence the 10's complement of 3764 is 6236. Adding 10's complement to 5129 produces 11365. Dropping the most significant digit 1, we get the answer as 1365.

Once you understand the algorithm, the working can be organized as follows :

9's complement of 3764=6235 (This step is oral.)
10's complement of 3764=6236 (Oral.)

```
 5129                          5129
-3764←10's complement→      +6236
 ----                      ------
                          (1) 1365   temporary sum
                          --------
                              1365   final sum
                                     (=desired difference).
```

Thus the answer is 1365.

Example 6. *Subtract 2371 from 9612 using 9's complements.*

Solution.

```
 9612                     9612
-2371←9's complement→   +7628
 ----                  ------
                    (1) 7240   temporary sum
                    +      1   end-around carry
                    --------
                        7241   final sum
```

7241 is the answer.

Example 7. *Subtract 3421 from 51192 using 10's complements.*

Solution. This problem is different from the others solved so far in as much as the sizes of the minuend and the subtrahend are different. One has five digits and the other only four. For the above method to work, the sizes of the two numbers should be same If we wish to allow different sizes, then we must modify the algorithm a little. Our most significant digit in this case need not be the 1. Thus instead of 'dropping the most significant digit', we shall have to subtract 1 from it. Let us demonstrate these remarks.

(*i*) Writing the subtrahend so that it has the same size as the minuend, we have

3421=03421.

9's complement of 03421 is 96578.
10's complement of 03421 is 96579.

```
 51192                        51192
-03421←10's complement→     +96579
 -----                     -------
                          (1)47771   temporary sum
                          --------
                             47771   final sum
```

Thus 51192−3421=47771.

```
(ii) 51192                          51192
   − 3421←10's complement→+  6579
   ------                   ------
                          (5) 7771   temporary sum
                            ------
                             47771   final sum.
```

Remark. If the two numbers are not the same size, a similar adjustment must be made while using the 9's complement. The best policy is to make both the numbers the same size.

Having understood how a computer performs a subtraction by turning it into an addition sum, let us have some practice in doing subtraction problems by hand on floating point representations of real numbers. No new concepts are involved and the familiar rules prevail. As before, we shall keep four digits for mantissa and two for the exponent, signs excluded. The only problem which may arise is that of *underflow*. If the difference is numerically less than ·1000 E−99, *i.e.*, if it falls in the range]−·1000 E−99, ·1000 E−99[, we cannot store it in the machine and the outcome of trying to solve such a problem on our compuler would be to receive an error message regarding this condition of underflow.

Example 8. *Subtract ·3721 E−05 from ·6912 E−05.*

Solution. The minuend is greater than the subtrahend. Therefore, the answer would be positive. The exponents are equal. Thus the working is straight-forward.

```
 ·6912 E−05
−·3721 E−05
-----------
 ·3191 E−05    Ans.
```

Example 9. *Subtract ·6124 E 05 from ·1109 E 06.*

Solution. Exponents are different. As in addition sums, we shall keep the term with higher exponent intact and adjust the other one so that this too has the same exponent.

```
 ·1109 E+06      ·1109 E+06
             ↔
−·6124 E+05     −·0612 E+06
-----------     -----------
                 ·0497 E+06=·4970 E+05    Ans.
```

Notice that the difference was not in the normalized from. By scaling, the final answer is given in the normalized form.

Example 10. *Subtract −·3795 E−5 from ·5795 E 1.*

Solution. Since subtracting a from b is equivalent to adding $-a$ to b, this problem is equivalent to the addition sum ·5795 E 1 +·3795 E−5.

Example 11. *Subtract ·3764 E 2 from ·1595 E 2.*

Solution. Here the subtrahend is greater than the minuend. Hence we first subtract ·1596 E 2 from ·3764 E 2 and take the negative of the difference as the answer. In other words, to evaluate the expression (·1596 E 2−·3764 E 2) we evaluate the equivalent quantity−(·3764 E 2−·1596 E 2), getting the answer−·2168 E 2.

Example 12. *Subtract−·2614 E+97 from−·3611 E+97.*

Solution.

$$-{\cdot}3611\text{ E}+97-(-{\cdot}2614\text{ E}+97)=-{\cdot}3611\text{ E}+97+{\cdot}2614\text{ E }97$$
$$=-({\cdot}3611\text{ E}+97-{\cdot}2614\text{ E}+97)$$

etc.

EXERCISE 15 (*c*)

1. Write the 9's complement :

   (*a*) 376, (*b*) 1078,
   (*c*) 23456, (*d*) 987654,
   (*e*) 0, (*f*) 99,
   (*g*) 0541.

2. Write the 10's complement :

   (*a*) 267, (*b*) 7180,
   (*c*) 654321. (*d*) 999999.

3. Write the 1's complement :

   (*a*) 110011, (*b*) 10101,
   (*c*) 100001, (*d*) 111111.

4. Write the 2's complement of each number in the previous problem.

5. Using 9's complements, subtract :

   (*a*) 156 from 237, (*b*) 3297 from 4556,
   (*c*) 37221 from 65000, (*d*) 13212 from 57621,
   (*e*) 26 from 350, (*f*) 181 from 2900.

6. Verify the subtractions in problem 5 above by using 10's complements.

7. Evaluate using 2's complements and verify by using 1's complements :

   (*a*) 111000−110011, (*b*) 1010101−1001001,
   (*c*) 11100−1000, (*d*) 101111−1011.

8. Subtract (and put the answer in the normalized floating point form) :

   (*a*) ·3740 E+03 from ·7912 E+03,
   (*b*) ·2711 E+06 from ·9700 E+06,

(c) ·4234 E+05 from ·1239 E+06,
(d) ·9712 E−31 from ·8876 E−30'
(e) ·3627 E+15 from ·4138 E+15,
(f) ·1939 E+98 from ·1143 E+99,
(g) ·2122 E+92 from ·4432 E+97.

9. Evaluate ;
(a) ·2537 E+21−·1224 E+14,
(b) −·2691 E+23+·3718 E+23,
(c) −·5990 E+44−·6000 E+47,
(d) ·6187 E+21−·9985 E+21.

10. Which of the following result in an underflow ?
(a) ·3411 E+99−·3400 E+99.
(b) ·5589 E+99−·5587 E+99.
(c) ·3333 E−99−·3000 E−99.

15·4·3. Multiplication and Division with Floating Point Representations of Real Numbers

The processes of multiplication and division are different than those of addition and subtraction in the arithmetic of floating point representations in one respect. Because of the rules

$$(a\times 10^m)\,(b\times 10^n)=(ab)\,10^{m+n},$$

and $$(a\times 10^m)\div(b\times 10^n)=\left(\frac{a}{b}\right)10^{m-n},$$

the problem of making the exponents same does not present itself here. Assuming the normalized forms, we have the following rules for multiplication and division :

Multiplication. The mantissa of the product of floating point representations of two real numbers is obtained by multiplying the mantissas of the two numbers. Only the four most significant digits are retained. The exponent of the product is the sum of the two exponents.

Division. The mantissa of the quotient of a floating point representation of a real number by another is obtained by dividing the mantissa of the first by that of the second. The exponent is obtained by subtracting the exponent of the denominator from that of the numerator.

Example 13. *Multiply 1237·5×10³ and ·094×10⁶ using normalized floating point representations with a three-digit mantissa and two-digit exponent.*

Solution. The given numbers written in the specified normalized form are ·123 E+07 and ·940 E+0·5. We have dropped

the two least significant digits in the first number in order to keep a 3-digit mantissa.

Now $(\cdot 123 \text{ E}+07)\times(\cdot 940 \text{ E}+05)=(\cdot 123\times\cdot 940) \text{ E}+(7+5)$,

$$=\cdot 115620 \text{ E}+12=\cdot 11562 \text{ E}+12.$$

Truncating (chopping off) the mantissa after three digits, the required answer is $\cdot 115 \text{ E}+12$.

Example 14. *Find the normalized floating point representation of 24576 and 81·4728 with a 5-digit mantissa and a 1-digit exponent. Divide the first number by the second and express the answer in the normalized form.*

Solution. $24576=\cdot 24576\times 10^5=\cdot 24576 \text{ E}+5.$

$81\cdot 4728=\cdot 814728\times 10^2=\cdot 81472 \text{ E}+2$, dropping the least significant digit (8) in order to keep a 5-digit mantissa. Hence

$$\frac{24576}{81\cdot 4728}=\frac{\cdot 24576 \text{ E}+5}{\cdot 81472 \text{ E}+2}=\frac{24576}{81472}\text{ E}+(5-2),$$

$$=\cdot 3016496 \text{ E}+3,$$

$$=\cdot 30164 \text{ E}+3, \text{ after chopping off 9 and 6.}$$

Remark. When two normalized floating point numbers are being multiplied, the mantissa of the product is always less than one (why ?). Hence there would never be an overflow in the mantissa. However, assuming a 4-digit mantissa, there could be 7 or 8 significant digits in the mantissa of the product (why ?). Hence three or four least significant digits must be chopped off in order to keep a 4-digit mantissa. In case of division, the mantissa of the quotient may exceed 1 (produce an example !). If so, scaling must be done in order to make the mantissa a fraction.

Example 15. *Divide ·3417 E−56 by ·1717 E−52 (and express the answer in the normalized floating point form).*

Solution. $(\cdot 3417 \text{ E}-56)\div(\cdot 1717 \text{ E}-52)=(3417\div 1717) \text{ E}(-56+52)$,

$$=1\cdot 990099 \text{ E}-04,$$

$$=\cdot 1990 \text{ E}-03.$$

Ans.

Here, to put the answer in the desired form, both scaling and chopping off have been done.

EXERCISE 15 (*d*)

Express all answers in the normalized floating point form.

1. Find the product in the normalized floating point form with a 3-digit mantissa and 1-digit exponent, mentioning the conditions of underflow or overflow whenever encountered :

(a) ·372 E 5×·169 E 2. (b) ·16 E−5×·798 E 9.
(c) ·612 E 9×·111 E−9. (d) ·274 E 8×·101 E 7.
(e) ·372 E−5×·193 E−6. (f) ·678 E 2×·403 E−2.

2. Which of the answers in problem 1 above would change and how on the assumption of a five-digit mantissa and a two digit exponent ?

3. Assuming a 4-digit mantissa and 2-digit exponent, divide
(a) ·3721 E 6 by ·2681 E 8. (b) ·2678 E 14 by ·1111 E 07.
(c) ·1000 E+17 by ·3345 E−09 (d) ·9999 E 99 by ·1111 E 11.

4. Do the sums in problem 3 above if the mantissa can hold 6 digits and the exponent 1. Which, if any, produces an overflow/underflow condition ?

5. Simplify the following and express the answer in the same type of floating point representations :
(a) ·379 E+09−·212 E+09+·111 E+10.
(b) −·2461 E+8+·3791 E+8−·1400 E+7.
(c) ·3951 E+06×·4221 E+03−·5555 E+10.
(d) ·51112 E+17÷·61122E+15+·99999 E+02.

6. Using a 4-digit fraction (*i.e.*, mantissa) and a 2-digit exponent, compute :
(a) $15\frac{1}{2}$% of 25941.
(b) The cost of 395 kg of sugar at Rs. 8.75 per kg.
(c) The cost of $695\frac{1}{4}$ m of tapestry at Rs. 68.95 per m.
(d) The mass of 5485 m^3 at 379 kg/m^3.
(e) The distance travelled in 9 *h* 10 m 11 sec at the speed of ·0005 km/sec.

15·5. INPUT ERRORS

The concept of approximations is not new to you. Eversince you started using the ruler as a *measuring device*, you have been *approximating*. Whether you measure to the nearest centimetre or millimetre, you know that there are lengths which cannot be measured exactly as so many centimetres or so many millimetres. The best you can do is to say *roughly* so many cm or *nearly* so many mm etc. Thus the measurements you have been working with need not be exact. This is true of most of the real life data too. Our data are hardly ever exact. Moreover, even when our data are exact, there is no guarantee that the tools with which we are going to process our data would produce exact results. For example, it is one thing to know that the length of the hypotenuse of a right triangle with each leg equal to one is $\sqrt{2}$; it is quite the fish of another pond to be able to *measure* this length and say *it is* $\sqrt{2}$. You can never obtain an output like $\sqrt{2}$ from a computer no matter how

many times you feed (1^2+1^2) and ask for the square-root. The reason need not be explained. You already know that a computer can store data in a finite form only. For example, with a six digit mantissa, a number like 12·34567 (or with more significant digits) cannot be stored exactly in a computer. For reasons such as these— *inexact data*, or *limitations of* computing *devices*, our results contain errors. Let us first examine the ways in which we *cut down* our data to represent them in a suitably chosen size. Then we shall consider the question of errors generated by such curtailing of numbers

15·5·1. Rounding of Numbers

As you know, the digits used to give an idea of the size of a given number apart from its exponential part are known as *significant figures* or *significant digits*, *e g.*, each of the numbers 12340, 1234, ·1234, 1·234, ·01234 has four significant figures. The zero in the first and the last of these numbers serves only to fix the position of the decimal point, and is not considered as a significant digit. However, a zero such as in 1023 *is* a significant figure. The sequence of significant digits in a number starts and ends with a non-zero digit. Sometimes it is necessary to cut down some of the digits of a given number because of the limitations of the framework we are working in. Since higher the positional value of a digit in the number, the greater the information carried by it, the digit in the leftmost (resp. rightmost) position is known as the *most significant* (resp. *least significant*) digit. Thus in 123, 1 is the most significant digit; 2 is the 2nd most significant digit, 3 is the least significant digit. Clearly, in order to lose as little information as possible, it is the least significant digits which should be cut off. The process of cutting down the digits is known as **rounding off**.

Let us illustrate the rules for rounding off through rounding off the numbers 346411, 126621, 686500, 565500 to three significant digits. The three most significant digits of these numbers are 346, 126, 686 and 565 respectively. The first step is to replace the remaining digits by zeros. This produces 346000, 126000, 686000 and 565000 respectively. The second step consists in modifying the *3rd* most significant digit because we wish to round off to *three* significant digits. The place value of 1 in the 3rd most significant figure here is 10^3 because it is the 4th from the right. In case of the first number 34640, the portion replaced by zeros is 411. This is less than *half the place value of a unit in the 3rd most significant place*, *i.e.*, it is less than $\frac{1}{2}\times 1000$ or 500. So we leave the *3rd digit unaltered* and get 346000 as the required rounded off value. In case of the second number 126621, the portion replaced by zeros is 621 which is greater than 500. So we *increase the 3rd digit by 1* and get 127000 as the rounded off value. In case of the third and the fourth numbers, the portion replaced by zeros is exactly 500. The third digit in case of 686500 is even and is, therefore, left unaltered.

In case of 565500, the third digit is odd and is, therefore, increased by 1. The rounded off numbers are 686000 and 566000 respectively.

We now state the rules for rounding off a number formally. *To round off a number to K significant figures*, the following rules are adopted :

1. To round off a whole number to K significant digits,

(*i*) replace all digits to the right of the Kth most significant digit by zero ;

(*ii*) modify the Kth most significant digit according to the following rules :

If the portion replaced by zeros is

(*a*) less than half the place value of a unit in the Kth most significant place, then keep it unaltered.

(*b*) greater than half the place value of a unit in the Kth place, increase the digit in the Kth place by 1.

(*c*) equal to half the place value of a unit in the Kth place, then leave the digit in the Kth place unaltered or increase it by 1 according as the digit in the Kth place is even or odd.

2. To round off a decimal fraction to K places of decimal,

(*i*) Chop off all the digits to the right of the Kth decimal place.

(*ii*) Increase the digit in the Kth decimal place by 1 if either the place value of the chopped off number is greater than half the place value of a unit in the Kth decimal place, or if these two are equal but the digit in the Kth place is odd. Otherwise leave the digit in the Kth decimal place unaltered.

Example 16. *Round off the following numbers to 4 significant figures* :

(*i*) *12346.* (*ii*) *345628.* (*iii*) *3412500,* (*iv*) *8713500.*

Solution. (*i*) The 4th most significant digit is 4. Replacing all the digits to the right of 4, we get 12340. The place value of a unit in the 4th most significant digit is 10 because this 4th digit occupies the second place from the right. Half of this value is 5. The number replaced by zero is 6. Since 6 is greater than 5, we increase the digit in the 4th place by 1. This gives us the rounded off number as 12350.

(*ii*) The 4th most significant digit in 345628 is 6 and the place value of a unit in this place is 100. Replacing all the digits to the right of 6, we get 345600. The number replaced by zeros is 28. Also, $28 < \frac{1}{2} \times 100$. Hence we leave the 4th digit, viz, 6, unchanged. Thus 345628 rounded off to four significant figures is 345600.

(*iii*) The 4th most significant digit is 2 and its place value is 10^3. The portion to the right of 2 is 500. Replace this 500 by 000 and note that 500 is equal to $\frac{1}{2}\times 10^3$. Since the digit 2 in the 4th place is even, it is to be left unaltered. Thus the rounded off number is 3412000.

(*iv*) The 4th most significant digit 3 is odd. The portion 500 to the right of this 3 is exactly half the place value of a unit in the 4th place. Hence increasing 3 by 1 and replacing 500 by 000, we get the required rounded off number as 871400.

Example 17. *Round off 3·14257 to* (*i*) *two places of decimal,* (*ii*) *three places of decimal.*

Solution. (*i*) The digit in the second decimal place is 4. The actual value of the digits 257 to be chopped off* is ·00257. The place value of a unit in the second decimal place is 10^{-2} or ·01. Half of ·01 is ·005. Also, ·00257 < ·005. Hence 3·14257 rounded off two decimal places is 3.14.

(*ii*) The third decimal place is occupied by 2. The value of the chopped off number is ·00057. The place value of a unit in the third decimal place is ·001 and half of this number is ·0005, which is less than the chopped off number ·00057. Hence increasing the digit in the third decimal place by 1 and chopping off the extreme 57, we get the rounded off value as 3·143.

Example 18. *Round off each of 9·075 and 9·065 to two places of decimal.*

Solution. In each case, the digit to be chopped off is 5 and the number which this 5 represents is ·005. The place value of a unit in the second decimal place is ·01. Half of ·01 is ·005 ; and this equals the chopped off number. Since the digit 7 in the second decimal place of 9·075 is odd, increasing it by 1, the rounded off value of 9·075 is 9·08. Since 6, the digit in the second decimal place of 9·065 is even, 9·065 rounded off to two decimal places is 9·06.

Note. Decimal places are counted from the right of the decimal point no matter what the digits occupying these places. Thus the 2nd decimal digit of 9·065 is 6 and not 5.

15·5·2. Truncating Numbers

The word meaning of *truncate* is to *cut short, to lop*. When we chopped off numbers to keep only four digits in the mantissa, we were *truncating* the numbers at four digits. Thus that chopping off was a very simple type of truncation. A more usual and important type of truncation is encountered when we approximate numbers by means of an infinite series. For example, we have

*Replacing 257 by 000 in 3·14257 amounts to chopping off 257 from 3·14257.

$$e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\ \cdots\cdots$$

To calculate e^x from this series, we may ignore all terms from some stage onwards, taking the sum of the first few terms as the value of e^x. For example, if we ignore the remainder after $n+1$ terms, an approximate value of e^x is

$$1+x+\frac{x^2}{2!}+\cdots\cdots+\frac{x^n}{n!}.$$

This is known as *truncating the series at the* $(n+1)$th *term.* For example, truncating at the 10th term, an approximate value of e^3 is

$$1+3+\frac{3^2}{2!}+\cdots\cdots+\frac{3^9}{9!}.$$

Notice that chopping off is a special case of the above type of truncation, since a decimal fraction x_1x_2......is nothing but the series

$$\frac{x_1}{10}+\frac{x_2}{10^2}+\cdots\cdots+\frac{x_n}{10^n}+\cdots\cdots$$

EXERCISE 15 (*e*)

1. Round off each of the following numbers to four significant digits :
   23466, 24573, 112009, 125603, 12455, 12325, 126255, 126295, 126550, 126650, 12598, 42695, 42685.
2. Round off each of the following to three decimal places :
   3·7981, 4·5678, 2·0912, 2·9097, 2·9985, 2·6665, ·7455, ·84559, ·66675, ·55533.
3. Decide whether the following thumb-rule for rounding off to k decimal places is correct :

   "Look at the $(k+1)$th digit. If it is less than 5, then replace all digits to the right of the kth digit by zeros to get the rounded off number. If this digit is 5 and all the digits to its right are zero, then (*i*) make this also zero, and (*ii*) make the kth digit even by adding one to it if not already even. The number so obtained is the required approximation. In all other cases increase the kth digit by 1 and replace all the digits to the right of the kth digit by zeros to get the required approximation.

15·5·3. Classification of Input Errors

We have seen that due to limitations of either measurement or representation, our data or input may be inexact giving rise to errors. Errors may also creep in due to the operations we perform on our data. However, for the sake of simplicity, we are assuming at the moment that all the arithmetic operations we carry out on our data are *exact*. The errors are entirely in the input.

Error. If x is an approximation to a true value X, then $x-X$ is known as the **error in x as an approximation to X**, or simply as *error in x.*

Illustration 1. The value of 5·176 rounded off to two decimal places is 5·18. The error in 5·18 as an approximation to 5·176 is $5{\cdot}18-5{\cdot}176=0{\cdot}004$.

Absolute Error. If x is an approximation to the true value X, then $|x-X|$ is known as the **absolute error** in x. Thus if e_a denotes the absolute error, then

$$e_a=|\text{approximate value}-\text{true value}|.$$

Illustration 2. The value of 1321 rounded to 3 significant figures is 1320. The error in 1320 is $1320-1321$, or -1. The absolute error in 1320 is $|1320-1321|$, or 1.

Note that the absolute error e_a is always positive.

Relative Error. Can you say which error is more serious, an error of 1 g in xk of rice or an error of 10 g in y k of rice ? You are wise, if you do not want to label one of them as more serious without looking into what x and y are. If x is one and y is 100, certainly the error 1 g is more serious than 10 g. Thus (absolute) errors may not convey the true picture. Hence it is more useful to use the concept of **relative error** e_r defined as follows :

$$e_r=\left|\frac{x-X}{X}\right|,$$

where x is an approximation to the true value X. For example, if 30 is an approximate value of 27 in some sense, then the relative error in 30 is $\left|\frac{30-27}{27}\right|$, or 1/9. It shows that the error (3) in the approximate value is 1/9th of the true value (27).

Remark. $100e_r$ is known as the **percentage error.**

Truncation Errors. Errors caused due to truncation are known as **truncation errors.**

Rounding Errors. Errors caused due to the rounding off of numbers are known as **rounding errors.**

Errors in Illustrations 1 and 2 are as a matter of fact rounding errors.

Example 19. *Show that if a number is rounded to 3 decimal places, then the maximum absolute error is* $(1/2)\times 10^{-3}$ *or* ·0005.

Solution. Suppose that the given number is $a{\cdot}x_1x_2\ldots\ldots$, where a is the integer part and $\cdot x_1x_2\ldots\ldots$ is the fractional part. Notice that while subtracting the true value from the rounded off value, the integer part would vanish. So it is sufficient to consider

a number of the type $.x_1x_2\cdots\cdots$, where each x_i is a digit from 0 to 9. Let us write the given number as

$$.x_1x_2\cdots\cdots=.x_1x_2x_3+{\cdot}000x_4x_5x_6\cdots\cdots$$

Increasing x_3 by 1 amounts to adding ·001 to $\cdot x_1x_2x_3$. The portion to be replaced by zeros is $\cdot 000x_4x_5\cdots\cdots$ The place value of a unit in the third place is 10^{-3}. The rounded off value is either $\cdot x_1x_2x_3$ or $\cdot x_1x_2x_3+{\cdot}001$. Also our rules for rounding off are such that when the rounded off number is $\cdot x_1x_2x_3$, then $\cdot 000x_4x_5\cdots\cdots \leqslant 1/2\times 10^{-3}$ or ·0005, and when the rounded off number is $\cdot x_1x_2x_3+{\cdot}001$, then $\cdot 000x_4x_5 \geqslant {\cdot}0005$. Thus when the absolute error,

$$|\text{ approximation} - \text{true value }|,$$

is given by

$$e=|\cdot x_1x_2x_3-(\cdot x_1x_2x_3+.000x_4x_5\cdots\cdots)|,$$

then $\qquad \cdot 000x_4x_5\ldots \leqslant {\cdot}0005,$

and when $\qquad e=|(\cdot x_1x_2x_3+{\cdot}001)-(\cdot x_1x_2x_3+{\cdot}000x_4x_5\cdots\cdots)|,$

then $\qquad \cdot 000x_4x_5\ldots\ldots \geqslant {\cdot}0005.$

Thus if $e={\cdot}000x_4x_5\cdots\cdots$, then we know that $\cdot 000x_4x_5\ldots\ldots \leqslant {\cdot}0005$ and when $e=|{\cdot}001-{\cdot}000x_4x_5\cdots\cdots|$, then we know that

$$\cdot 000x_4x_5\cdots\cdots \geqslant {\cdot}0005.$$

In the former case, $e \leqslant {\cdot}0005$, and in the latter,

$$\begin{aligned} e &= |{\cdot}001-{\cdot}000x_4x_5\ldots\ldots|, \text{ and } {\cdot}000x_4x_5\ldots\ldots \geqslant {\cdot}0005, \\ &= |{\cdot}001-({\cdot}0005+a)|, \text{ where } 0 \leqslant a < {\cdot}0005, \\ &= |{\cdot}0005-a|, \text{ where } 0 \leqslant a < {\cdot}0005 \\ &\leqslant {\cdot}0005. \end{aligned}$$

Hence in both the cases, $e \leqslant .0005$ or $(1/2)\times 10^{-3}$.

Remark. A similar argument may be used to prove that if a number is rounded to n places of decimal, then the absolute error is less than or equal to $(1/2)\times 10^{-n}$.

Error bounds. While calculating errors above, we tacitly assumed that we knew the actual value. In practice, however, we generally do not know the actual value. For example, when we measure lengths with the help of a ruler, we might label the lengths as 1, 2, 3, 4, ...cm. Lengths between 2 cm and 3 cm would be labelled either as 2 cm or 3 cm. We would not know the *exact* length but we know for certain that the absolute error in our measurement is less than 1 cm. This 1 cm is known as an **absolute error bound**. If x is an approximation to the true value X (unknown) and $|x-X| \leqslant \varepsilon$, then we say that ε is an error bound. The true value X then lies in the interval $[x-\varepsilon, x+\varepsilon]$ (We may similarly define **relative error bound**). $(1/2)\times 10^{-n}$ is an absolute error bound when we round off to n decimal places.

Example 20. *Find an absolute error bound for the mantissas of normalized floating point numbers when stored in a machine which truncates the mantissas at four digits.*

Solution. Here, the stored mantissas are the approximations. Just for the sake of understanding, notice that a mantissa like ·12345 is stored as ·1234. Are there any others mantissas which would be stored as ·1234 ? Since a stored mantissa is always less than or equal to the true mantissa, all such mantissas are greater than or equal to ·1234. Also, ·1235 and all bigger mantissas would be stored as atleast ·1235. All the mantissas in the interval [·1234, ·1235 [would be stored as ·1234. Hence the maximum absolute error is $|\,\cdot 1235 - \cdot 1234\,| = \cdot 0001 = 10^{-4}$. The same argument works no matter what the mantissa is. Hence 10^{-4} is a suitable error bound. (Which method gives better approximations—chopping off or rounding off ? Equivalently, which type of errors are smaller ?)

Remarks. **1**. If ϵ is an absolute error bound, so is every number greater than ϵ. However, when we talk about an error bound, we wish to go as close as possible. Thus 'bound' is generally used here in the sense of 'least upper bound'.

2. We can find an absolute relative error bound also in the above example. If the floating point representation of the given number is $\cdot x_1x_2x_3x_4\text{E}x_5x_6$, where each x_i is a digit, $x_1 \neq 0$, then as above, the maximum absolute relative error is

$$\frac{\cdot 0001\text{E}x_5x_6}{\cdot x_1x_2x_3x_4\text{E}x_5x_6} = \frac{1}{x_1x_2x_3x_4} \leqslant \frac{1}{1000},$$

because $x_1x_2x_3x_4 \geqslant 1000$. Hence 10^{-3} is an absolute relative error bound. Clearly, negative mantissas and exponents warrant no change in the argument.

Example 21. *If the normalized floating point mantissas are rounded off to four decimal places (keeping a four-digit mantissa), then determine a convenient relative absolute error bound.*

Solution. We know that the absolute maximum rounding off error in this case is $(1/2) \times 10^{-4}$ or $\cdot 5 \times 10^{-4}$. Hence denoting the absolute relative error by e_r,

$$e_r \leqslant \frac{\cdot 5 \times 10^{-4}}{\cdot x_1x_2x_3x_4x_5\cdots\cdots}$$

But $\qquad \cdot x_1x_2x_3x_4\cdots\cdots \geqslant \cdot 1, \qquad \because \quad x_1 \geqslant 1,$

$$\Rightarrow \frac{1}{\cdot x_1x_2x_3\cdots\cdots} \leqslant 10,$$

$$\Rightarrow \frac{\cdot 5 \times 10^{-4}}{\cdot x_1x_2x_3\cdots\cdots} \leqslant 10 \times \cdot 5 \times 10^{-4},$$

$$\Rightarrow e_r \leqslant 5 \times 10^{-4}.$$

Remark. From the above example, it follows that *in the normalized floating point representation with four-digit mantissa, the absolute rounding error is less than or equal* to 5×10^{-5} *and the absolute relative rounding error is less than or equal* to 5×10^{-4}.

EXERCISE 15 (*f*)

1. Approximate by chopping off the least significant digits keeping only 2 digits in the mantissa and find the absolute errors :
·37982, ·36451, ·132956, ·987600.

2. Round off to three decimal places and find the absolute rounding errors :
3·7984, 3·6548, 7·9125, 7·9135.
For which of these numbers, the errors are different from the corresponding absolute errors ?

3. Round off each of the following numbers to three significant digits :
35787, 26401, 31009, 4002.
Find
(*a*) the errors.
(*b*) the absolute errors.
(*c*) the relative absolute errors.
(*d*) the percentage errors.

15·6. ERROR ARITHMETIC

Let us consider to assume that the errors are due to input alone and that the arithmetic operations on our data are exact. Let us see what we can say about errors in sum, difference, product and quotient of two approximations when the errors or error bounds for them are known.

(*a*) *Error in the sum.* Let x_1, x_2 be approximations to X_1 and X_2 and e_1, e_2 respectively the errors therein, so that

$$x_1-X_1=e_1,\ x_2-X_2=e_2.$$

But then $(x_1-X_1)+(x_2-X_2)=e_1+e_2$,

or $(x_1+x_2)-(X_1+X_2)=e_1+e_2$.

Hence error in (x_1+x_2)=(error in x_1)+(error in x_2).

If ε_1, ε_2 are absolute error bounds instead of being the actual errors, then

$$|(x_1+x_2)-(X_1+X_2)| \leqslant |x_1-X_1|+|x_2-X_2| \leqslant \varepsilon_1+\varepsilon_2.$$

Hence an absolute error bound for the sum is $\varepsilon_1+\varepsilon_2$.

(*b*) *Error in the difference.* As before, the error e_2 in the difference x_1-x_2 is given by

$$e_2=(x_1-x_2)-(X_1-X_2),$$
$$=(x_1-X_1)-(x_2-X_2)=e_1-e_2.$$

Hence error in (x_1-x_2)=(error in x_1)−(error in x_2).

Similarly, denoting the absolute bounds for x_1, x_2 respectively by ε_1 and ε_2,

$$|(x_1-x_2)-(X_1-X_2)| \leqslant |x_1-X_1|+|x_2-X_2| \leqslant \varepsilon_1+\varepsilon_2.$$

Hence $\varepsilon_1+\varepsilon_2$ is an absolute error bound for x_1-x_2. Note that error bounds have been *added* even for difference of approximate numbers.

(*c*) *Errors in the product and quotient.* It can be shown that the relative error in the product (resp. quotient) of two approximate numbers is the sum (resp. difference) of the relative errors in the two numbers. If however, we consider absolute relative errors or error bounds, then these are added for both the product and the quotient of the approximate numbers.

Example 22. *How accurately should the sides of a room be measured so that the maximum error in computing the perimeter may be less than 20 cm.*

Solution. Since errors in a sum are added up, it would be safe to measure each side with an error bound of 5 cm.

Example 23. *The length of a metal bar at a certain temperature has been measured to be 30·32 cm with an error bound of $\frac{1}{2}\times 10^{-2}$ cm. Determine the range in which the true length of the bar must lie.*

Solution. $\frac{1}{2}\times 10^{-2}=\cdot 005$. Since the approximate value can not be in an error more than ·005 cm, therefore, the true value (in cm) must lie in the interval [30·32−·005, 30·32+·005], or [30·315, 30·325]. In other words, 30·315 cm $\leqslant$ true length $\leqslant$ 30·325 cm.

Example 24. *Rounded off to 3 decimal places, the values of $\sqrt{5}$ and $\sqrt{7}$ are respectively 2·236 and 2·646. Estimate (i) the absolute error in calculating $\sqrt{5}+\sqrt{7}$ from these values, and (ii) the absolute relative error in calculating $\sqrt{35}$ from these values.*

Solution. (*i*) Since we are rounding off to three places of decimal, the rounding off error in each of $\sqrt{5}$ and $\sqrt{7}$ is $\leqslant \frac{1}{2}\times 10^{-3}$ (or ·0005). Hence the error in the sum is $\leqslant \frac{1}{2}\times 10^{-3}+\frac{1}{2}\times 10^{-3}$, or 10^{-3}.

(*ii*) The maximum error due to rounding is ·0005.

By (*c*) above, the relative error E satisfies

$$E \leqslant \frac{\cdot 0005}{\sqrt{5}}+\frac{\cdot 0005}{\sqrt{7}},$$

$$\leqslant \frac{\cdot 0005}{2\cdot 000}+\frac{\cdot 0005}{2\cdot 000}=\cdot 0005.$$

$(\because\ 2\cdot 236 \geqslant 2\cdot 000 \text{ etc.})$

EXERCISE 15 (*g*)

1. If $x=15\cdot 0$ and $y=7\cdot 5$ are approximations to the numbers $X=15\cdot 6$ and $Y=7\cdot 8$, then write down the error in each of the following approximations :

(*a*) $x+y$ as an approximation to $X+Y$.

(*b*) $x-y$ as an approximation to $X-Y$.

(*c*) xy as an approximation to XY.

(*d*) x/y as an approximation to X/Y.

(*e*) $x-3y$ as an approximation to $X-3Y$.

2. If error bounds for two approximate numbers are ·005 and ·001, then find an error bound for

(*a*) $x+y$, (*b*) $x-y$.

3. If relative errors in two approximate number x and y are ·0001 and ·05 respectively, then find the relative errors in the sum $x+y$ and product xy.

4. If you wish to measure x and y, and if you wish to compute the difference $x-y$ with an error bound 1 cm, how accurately should you measure x and y ?

15·7. NUMERICAL INSTABILITY

We have seen that errors are added when approximate numbers are added or multiplied. Thus if an intput $x=\cdot 1234$ is in an error of ·01 cm, then $100x$ would be in an error of 1 cm. 10^6x would contain an error of 1 m. The output would, therefore, be very unrealistic. This is a kind of numerical instability. By numerical instability *we mean that behaviour by reason of which a small error at some stage of computation gets magnified to produce a large error in the final answer*. There are two different types of instabilities. The first type is caused by the nature of the situation or the data regarding the problem at hand. The second type—the more serious —is caused by the method we adopt for the solution of the problem. Both of these will be discussed one by one.

15·7·1. Inherent Instability

When an input value x is not exactly known, and ε is an absolute error bound, then all we can say about the true value of x is that it lies in the interval $[x-\varepsilon, x+\varepsilon]$. In a practical (and even theoretical) problem, we might be interested in finally computing $f(x)$ for some function f. Depending upon the nature of f, the original error (*i*) may increase unreasonably, (*ii*) may remain moderate, or (*iii*) may even be reduced, as the following examples show. In the first case we call the problem *ill-conditioned* or *inherently instable*, as the reason for instability are inherent in the problem itself rather than having something to do with the methods of solution.

Example 25. *If the measurement of a certain quantity X correct to two significant figures is 3·00, what can you say about the value of* (*i*) X^8, (*ii*) $X^{1/20}$, (*iii*) $100X$?

Solution. (*i*) Correct to two significant figures here means correct to one decimal place. Hence the error bound due to rounding is $\frac{1}{2}\times 10^{-2}$ or $\cdot 05$. Hence the true value X lies in

$$[3\cdot 00-\cdot 05,\ 3\cdot 00+\cdot 05] \text{ or } [2\cdot 95,\ 3\cdot 05].$$

Now $X \in [2\cdot 95, 3\cdot 05] \Rightarrow X^8 \in [2\cdot 95^8, 3\cdot 05^8]=[5736, 7489]$ nearly. Had we used the approximate value 3 of X, we would have obtained an approximation 3^8 or 6561. Since the true value of X^8 is some value between 5736 and 7489, even the two most significant digits in the approximate value of X^8 could be both wrong.

(*ii*) As in the previous case, we find that

$$X^{1/20} \in [2\cdot 95^{1/20}, 3\cdot 05^{1/20}]=[1\cdot 055, 1\cdot 057] \text{ nearly.}$$

Also, $3^{1/20}=1\cdot 056$. Thus had we used the approximate value 3 of X to evaluate $X^{1/20}$, the three most significant digits would have been certainly correct.

(*iii*) $100X \in [100\times 2\cdot 95, 100\times 3\cdot 05]=[295, 305]$.

Also, $100\times 3=300$, showing that like the input, the output is correct to two significant figures.

Remarks. 1. Problem (*i*) above is *ill-conditioned* because the errors in data are enlarged but (*ii*) is *well-conditioned*, because the errors in data get reduced here.

2. Accuracy in the output may increase, decrease or remain unaltered as in case of (*ii*), (*i*) and (*iii*) respectively.

15·7·2. Induced Instability

As you have seen, all computer arithmetic is carried out in floating point representations. The floating point operations are never exact due to constant adjustments of the mantissa and the scaling (adjustment of the exponent). For the sake of simplicity, we had assumed that the numbers are chopped off after a certain number of significant digits. As a matter of fact, most machines *round off* rather than *chop off*. This reduces error to a small degree, the errors nevertheless are there. Most of our output are obtained through the four fundamental arithmetic operations plus the extraction of roots (square-roots, cube-roots, ...). Sometimes several different sequences of these operations may be used to evaluate the same expression. For example, $(x+y)^2$ may be evaluated as $(x+y)(x+y)$, or as $x^2+2xy+y^2$, or as $y^2+2xy+x^2$ or as $x(x+y)+y(x+y)$ and so on. Believe it or not but different ways of evaluating the same expression may produce different results as we shall soon demonstrate. The output obtained in one way may have less error than that obtained from using another algorithm. This is where the question of *good methods versus bad*, pops up. Sometimes it is possible to avoid the unreasonable error which arises from using a particular algorithm by the use of a different algorithm. Under the circumstances, the instability caused by using a *bad method* is

known as the **induced stability** ; it is *induced* by the method used rather than being inherent in the problem.

Example 26. *Sum $3{\cdot}45+0{\cdot}0099+0{\cdot}0023$ using three-digit normalized mantissas. Find the error in the sum. Can this be reduced ?*

Solution. The true sum is $3{\cdot}4622$. Using three-digit mantissas,

$$\begin{aligned}3{\cdot}45+0{\cdot}0099+0{\cdot}0023 &= {\cdot}345\,E1+{\cdot}99E-2+{\cdot}23\,E-2,\\ &=({\cdot}345\,E1+{\cdot}99E-2)+{\cdot}23\,E-2,\\ &=({\cdot}345\,E1+{\cdot}000\,E1)+{\cdot}23\,E-2,\\ &={\cdot}345\,E1+{\cdot}000\,E1,\\ &={\cdot}345\,E1=3{\cdot}45.\end{aligned}$$

The error$=3{\cdot}45-3{\cdot}4622=-{\cdot}0122$.

Let us now find the sum in the reverse order.

$$\begin{aligned}0{\cdot}0023+0{\cdot}0099+3{\cdot}45 &=({\cdot}23\,E-2+{\cdot}99\,E-2)+3{\cdot}45,\\ &=1{\cdot}22\,E-2+3{\cdot}45,\\ &={\cdot}122\,E-1+{\cdot}345\,E1,\\ &={\cdot}001\,E1+{\cdot}345\,E1,\\ &={\cdot}346\,E1=3{\cdot}46.\end{aligned}$$

This time, the error$=3{\cdot}46-3{\cdot}4622=-{\cdot}0022$.

Thus the error has reduced.

Whenever there is an appreciable difference in the magnitudes of the numbers to be summed up, *they should be added in the ascending order of magnitude.* Why ? When we add using floating point addition, we raise the smaller exponent to make it equal to the bigger one. Raising the exponent means multiplying the number by a positive power of 10. To keep the number the same, we divide the mantissa by the same power of 10. Dividing by 10^k here means introducing k zeros after the decimal point which results in the loss of k digits from the mantissa. For example, if $k>4$, all the mantissa digits would be lost, and you would be virtually adding zero to the bigger number. Thus we would better add the smaller numbers first so as to make the sum as big as possible before we add a relatively big term.

Remark. The associative law of addition fails in the floating point arithmetic.

Example 27. *Use two different algorithms to find the third side of a right-angled triangle whose hypotenuse and one side are given to be $7{\cdot}5$ and $7{\cdot}2$. Assume that computations are done on a machine which rounds off to two significant figures. Compare the results.*

Solution. The third side is $\sqrt{(7{\cdot}5^2-7{\cdot}2^2)}$. Now

$$7{\cdot}5^2=56{\cdot}25=56 \text{ (to 2 significant figures).}$$

$7{\cdot}2^2=51{\cdot}84=52$ („ „).

$\therefore$ $7{\cdot}5^2-7{\cdot}2^2=4.$

$\therefore$ $\sqrt{(7{\cdot}5^2-7{\cdot}2^2)}=\sqrt{4}=2.$

The true value is 2·1. Hence there is an error $-{\cdot}1$.

We can use the formula $(a+b)(a-b)=a^2-b^2$ and compute the third side as $\sqrt{\{(7{\cdot}5+7{\cdot}2)(7{\cdot}5-7{\cdot}2)\}}=\sqrt{\{(14{\cdot}7)({\cdot}3)\}}=\sqrt{\{(15)({\cdot}3)\}}$, rounding off to two significant figures. Thus the third side$=\sqrt{4{\cdot}5}$ $=2{\cdot}12=2{\cdot}1$ which is the correct value. Hence the latter algorithm is better.

Remark. The latter algorithm was better because it involved only one multiplication as against the two ($7{\cdot}5\times7{\cdot}5$ etc.) of the former. Multiplication is repeated addition and every time you carry-out an addition, there is some chance of the sum of mantissas being greater than one. Whenever this happens, scaling is done to normalize the number, and 1 significant digit might be lost.

Example 28. *Evaluate the smaller root of $x^2-22x+1=0$ to our significant figures in two different ways. Are the values different ?*

Solution. Using the formula, the smaller root x_1 is given by

$$x_1=\frac{22-\sqrt{(22^2-4)}}{2},$$
$$=11-\sqrt{(120)},$$
$$=11-10{\cdot}9544=11-10{\cdot}95 \text{ to four significant figures},$$
$$=0{\cdot}05.$$

The bigger root $x_2=11+10{\cdot}95=21{\cdot}95$.

Product $x_1x_2=1$. Hence $x_1=1/x_2={\cdot}045558={\cdot}04556$.

Thus the latter method is better. The earlier algorithm (the formula) gives the answer to only two significant figures whereas the latter one gives accuracy to four significant figures.

Why did the formula method fail us here ? The reason is that we got involved in a difference of two almost equal numbers (11 and 10·9544). Notice that for almost equal numbers, the exponents would be the same and the mantissas almost equal. Hence the difference in the mantissas is a rather small number and results in some zeros after the decimal point. Since mantissa can contain only four digits, the least significant digits, which in this case carry the maximum information about the number, are lost.

All the above examples demonstrate one theme, viz., *some algorithms may be better than others in that they reduce induced instability*. Some usual causes of induced instability are :

(*a*) Subtraction of nearly equal numbers.

(*b*) Adding in descending order.

(*c*) Lengthy computations.

These should be avoided as far as possible. The accuracy in the output is determined not only by the degree of accuracy of the computational device in question but also by the method which we use to solve our problem.

EXERCISE 15 (*h*)

Note. *Carry out all operations in normalized floating point form with a four-digit mantissa and a 2-digit exponent unless stated otherwise.*

1. Add once in ascending and once in descending order of magnitude :

   (*a*) 47·961, 3·124, and 1·569.
   (*b*) 3827, 12·54, and 1·567.

   Compute the true value in each case. Which algorithm gives a result nearer the true value ?

2. Evaluate $(1/15)-(1/16)$, once as a difference of 1/15 and 1/16, and once as $1/(15\times 16)$. Which algorithm gives more significant digit in the output, and is hence better ?

3. Evaluate $9\times{}$·9999, once as $(10\times{}$·9999$-{}$·9999$)$ and once as a repeated addition of ·9999 with itself nine times. Find the true value. Which output is less in error ?

   [Multiplication by 10 is carried out by shifting the decimal point right. *Do not forget to normalize the output after each operation.*]

4. Evaluate $\sqrt{40}-\sqrt{39}$ once by taking the difference of the two roots and once as $1/(\sqrt{40}+\sqrt{39})$. Which output has more significant figures and is hence better ?

5. Evaluate the shorter root of each of the following equations in two ways after the fashion of example 28 in the text. Which method gives a better value ?

   (*a*) $x^2-18x+1=0$. (*b*) $x^2-24x+1=0$.
   (*c*) $x^2-2x+{}$·$009=0$. (*d*) $x^2-100x+1=0$.

6. Computations were carried out to 7 significant digits to solve the following pairs of equations :

   (A) $0·999999x+y=7,\ x+y=5.$ (B) $1·00001x+y=7,\ x+y=5.$

   The solution for (A) was $x=-1973790$ and $y=1973800$, and that for (B), $x=209715$ and $y=209710$.

   The instability is induced or inherent ?

   [**Hint.** The vast difference in the solution is caused by a small change in the coefficient of x in the first equation.]

15·8. ITERATIVE METHODS

Recall how you tried to obtain a decimal representation for $\sqrt{2}$. You noticed that $1^2=1$, $2^2=4$, and $(\sqrt{2})^2=2$ lies between 1 and 4. Thus you guessed the value of $\sqrt{2}$ to lies between 1 and 2. You then divided the interval] 1, 2 [into ten equal parts by means of the points 1·1, 1·2, 1·3,......, 1·9. You squared these numbers and noticed that $1{\cdot}1^2$, $1{\cdot}2^2$, $1{\cdot}3^2$, $1{\cdot}4^2$ were less than $(\sqrt{2})^2$ but $1{\cdot}5^2$ was greater than $(\sqrt{2})^2$. You deduced that $\sqrt{2}$ must, therefore, lie between 1·4 and 1·5. (This narrowed the range)] 1, 2 [to] 1·4, 1·5 [telling you that the first significant digit in $\sqrt{2}$ is 1). You then *repeated* the process of dividing the interval] 1·4, 1·5 [into ten parts by means of the numbers 1·41, 1·42, 1·43,......, 1.49, and then locating the interval] 1·41, 1·42 [which contained $\sqrt{2}$. This enabled you to say that $\sqrt{2}$ truncated to two digits is 1·4. This process was *repeated* or *iterated* again and again to get the various truncated values of $\sqrt{2}$ as 1, 1·4, 1·41, 1·414, 1·4142,...Such techniques of repeating, *i.e.*, *iterating* a process to get better and better solutions at each step are known as *iterative methods*. The three essential components of an iterative method are the following :

(*a*) *An initial approximation for the solution*, *e.g.*, in the above example, the first approximation to $\sqrt{2}$ was 1·0, since all values between 1 and 2 start with 1....... You must make a logical guess. Generally, it is not a difficult affair.

(*b*) *A rule for generating a new value* from the previous one and we hope the new value to be a better approximation than the previous one In the above example, the rule was to divide the previous interval into ten sub-intervals and picking up that one whose end-points a, b satisfied $a^2 \leqslant 2 \leqslant b^2$. This every time added one more significant digit in the value of $\sqrt{2}$ and hence gave a better approximation.

(*c*) An instruction about *when to stop or terminate the process*. In the above example, the rule could be : *stop when the square of a point of division becomes equal to* 2. That would give us the value of $\sqrt{2}$. But a *doubter* or a *lazy person* might have made you decide on another kind of termination rule. "*What if we never obtained a division point whose square is two ?*" "*How long do I go on ? May be, $\sqrt{2}$ has 20 decimal digits in it ! Thank you and sorry ! I do not have that much patience.*" Such questions/remarks as these are valid, really ! An iterative process may *never end* for all you know. Would the process in the above example ever end ? No ! On the other had our number been, 1/16 say, the process would have soon stopped. In fact it is not uncommon to find that after a certain stage, the new values cease to be better approximations. It would be futile to go on and on in such a case. In general, a few iterations may suffice. For example, six significant digits may be all we might need for using $\sqrt{2}$. When after a few iterations, the solution starts *settling down* or starts approaching nearer and nearer a certain

fixed value, we say that the process is *convergent*. A much-used rule in this connection is to take a small number ε in advance. When two successive approximations u, v are found such that $| u-u | <\varepsilon$, then we stop. The most common values ε of are 10^{-4} and 10^{-6}. Approximations u, v satisfying $| u-v | <\varepsilon$ are said to be *equal* in this context. Another way of expressing the same thing is to say that two successive approximations agree to a given number of significant digits.

We shall now demonstrate an iterative method to calculate the values of e^x for given values of x.

15·8·1. Evaluation of e^x

You know that

$$e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots+\frac{x^n}{n!}+\ldots\ldots \qquad \ldots(A)$$

$$=\frac{x^0}{0!}+\frac{x^1}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots+\frac{x^n}{n!}+\ldots \qquad \ldots(B)$$

Given a value of x, 5 say, is it possible to evaluate e^5 from the above series. Clearly, since the series never ends, it is impossible to substitute $x=5$ in the R.H.S. of (A) and evaluate the resulting expression in the usual sense of a sum. However, we notice that the general term T_n or $x^n/n!$ would be

$$T_n=\frac{5\times5\times5\times\ldots\ldots\ldots n \text{ times}}{1\times2\times\cdots\times5\times6\times7\times\ldots\times n}.$$

Notice that as n increases, numerator does not increase as fast as the denominator. The net effect is that as n increases, T_n decreases, It soon becomes less than one-tenth and thereafter the number of zeros after the decimal point goes on increasing. In any practical situation, we want accuracy upto a certain finite number of significant digits only. If we wanted our answer correct to four decimal places, then once we start getting terms like $\cdot 00000x_1x_2\cdots\cdots\cdots$, it is no use adding them ; they would contribute nothing to the sum. Two successive sums S, S′ would hence forth satisfy $| S-S' | <10^{-4}$. Thus by taking the sum of the first k terms for a suitable choice of k, we may evaluate e^x to the desired accuracy.

How would you calculate $x^2/2!$? A silly question ! You would multiply x by x and divide by 2. How else indeed ? What about $x^3/3!$ now ? Evaluate $x\times x\times x$, evaluate $3!$ and divide ? Indeed not ? We can economize on our computation. We have already evaluate $x^2/2!$ for the given value of x. All we need do is multiply it by $x/3$. This produces $(x^2/2!)\times x/3=x^3/3!$ Now we iterate this process. The next term, *i.e.*, the 4th term can be obtained from the 3rd by multiplying it with $x/4$.

In general

$$\frac{x^{n+1}}{(n+1)\,!}=\frac{x}{n+1}\times\frac{x^n}{n\,!},$$

or $$T_{n+1}=\frac{x}{n+1}\times T_n,$$

with $$T_0=1=\frac{x^0}{0\,!}.$$

Note : (*a*) The initial value is $T_0=1$. The first iteration produces T_1.

(*b*) Tha rule to get the T_n in the *n*th iteration is

$$T_n=\frac{x}{n}\times T_{n-1}.$$

(*c*) The process is to terminate when the first four digits after the decimal point are all zero. This means we shall chop off after four decimal places.)

Having understood the algorithm let us evaluate the terms in e^3 till we get five zeros following the decimal point. We shall use eight figure in our computation because most hand calculators use eight figure display. If you are using log-tables, or computing on paper and pencil, use lesser number of figures and lesser accuracy by all means.

Initializatian $I_0 : T_0=1$

First iteration $I_1 : T_1=\frac{x}{1}\ T_0=\frac{3}{1}\times 1,\ (x=3)$

$$=3.$$

Second iteration $I_2 : T_2=\frac{x}{2}\ T_1=\frac{3}{2}\times 3=4{\cdot}5$

and so on $$T_3=\frac{x}{3}\ T_2=\frac{3}{3}\times 4{\cdot}5=4{\cdot}5$$

$$T_4=\frac{x}{4}\ T_3=\frac{3}{4}\times 4{\cdot}5=3{\cdot}375$$

$$T_5=\frac{x}{5}\ T_4=\frac{3}{5}\times 3{\cdot}375=2{\cdot}025$$

$$T_6=\frac{x}{6}\ T_5=\frac{3}{6}\times 2{\cdot}025=1{\cdot}0125$$

$$T_7=\frac{3}{7}\ T_6=\frac{3}{7}\times 1{\cdot}0125={\cdot}4339285$$

$$T_8=\frac{3}{8}\ T_7=\frac{3}{8}\times{\cdot}4339285={\cdot}1627232$$

$$T_9=\frac{3}{9}\times{\cdot}1627232={\cdot}0542410$$

$$T_{10}=\frac{3}{10}\times{}^{\cdot}0542410={}^{\cdot}0162723$$

$$T_{11}=\frac{3}{11}\times{}^{\cdot}0162723={}^{\cdot}0044379$$

$$T_{12}=\frac{3}{12}\times{}^{\cdot}0044379={}^{\cdot}0011094$$

$$T_{13}=\frac{3}{13}\times{}^{\cdot}0011094={}^{\cdot}0002560$$

$$T_{14}=\frac{3}{14}\times{}^{\cdot}0002560={}^{\cdot}00000548$$

$$T_{15}=\frac{3}{15}\times{}^{\cdot}0000548={}^{\cdot}0000109$$

$$T_{16}=\frac{3}{16}\times{}^{\cdot}0000109={}^{\cdot}0000020.$$

Thus if we desire to turncate the sum at the 4th decimal place, there is no need to calculate the subsequent terms. The sum can also be found by using an iterative process. Denoting the sum of the first $n+1$ terms by S_n, we notice that

$$\begin{aligned}
S_0&=T_0\\
S_1&(=T_0+T_1)\\
&=S_0+T_1\\
S_2&=(T_0+T_1)+T_2\\
&=S_1+T_2\\
S_3&=S_2+T_3\\
&\vdots\\
S_n&=S_{n-1}+T_n\\
&\vdots
\end{aligned}$$

This time the iteration is to stop as soon as the first four digits of S_n and S_{n-1} become equal, because this will ensure that the first four digits after the decimal point have become zero and subsequent terms would contribute nothing to the sum so far as the accepted level of accuracy is concerned. The whole work can be organized in the form of a table as given on next page.

Let us explain the first two rows of the table. When $n=0$, $T_n=T_0=1$ (reading from the series) ;

$$T_{n+1}=T_1=\frac{3}{1}\times T_0=3\ ;$$

$$S_{n+1}=S_1=S_0+T_1=1+3=4.$$

n (Initialization)	nth term T_n ($T_0=1$)	(n+1)th term $T_{n+1}(=\frac{3}{n+1}\times T_n)$	Sum to (n+1) th term $S_{n+1}(=S_n+T_{n+1})$ ($S_0=1$)
0	1.0000000	$(\frac{3}{1}\times 1.0000000=)$3.0000000	(1+3=) 4.0000000
1	3.0000000	$(\frac{3}{2}\times 3.0000000=)$4.5000000	8.5000000
2	4.5000000	4.5000000	13.0000000
3	4.5000000	3.3750000	16.3750000
4	3.3750000	2.0250000	18.4000000
5	2.0250000	1.0125000	19.4125000
6	1.0125000	0.4339285	19.8464285
7	0.4339285	0.1627232	20.0091517
8	0.1627232	0.0542410	20.0633927
9	0.0542410	0.0162723	20.0796650
10	0.0162723	0.0044379	20.0841029
11	0.0044379	0.0011094	20.0852123
12	0.0011094	0.0002560	20.0854683
13	0.0002560	0.0000548	20.0855231
14	0.0000548	0.0000109	20.0855340
15	0.0000109	0.0000020	20.0855360

A good working strategy is to check the computed T_1 and S_1 directly from the series so that any mistakes made in the recurrence relation written for T_{n+1} and S_{n+1} may be rectified right here avoiding later misery. The series in question is $e^3=1+3+3^2/2\,!+3^3/3\,!+$......... We see that indeed $T_1=3$ and $S_1=4$. Now the second row. When $n=1$, $T_n=T_1=3$ (copying it down from the previous row) ; $T_{n+1}=T_2=(3/2)\times T_1=4{\cdot}5$; $S_{n+1}=S_2=S_1+T_2=4+4{\cdot}5=8{\cdot}5$, taking the value of S_1 from the last column of the previous row and T_2 just obtained. Verify directly from the series. You should carry on the verification on T_n now and then directly from the series in order to make sure you have not committed a mistake so far. Verifying S_n directly may not be possible at later stages.

The required sum turncated at 4th decimal place and correct to six significant digits is 20·0855. (Rounding off would also produce the same answer.)

Verification : An approximate value of $e=2{\cdot}71828$. Hence an approximate value of e^3 is 20·855, chopping off after four decimal places.

Are you thinking why did we go to all the trouble if it were so simple as all that ? If yes, try evaluating $e^{3\cdot 1}$ by this method ! The value 3 was chosen in the demonstration above for the sake of simplicity. The worth of the method comes to the fore when x is not an integer, and you do not have a hand calculator. Some of the normal functions can be evaluated directly (by the push of a few buttons) with the help of a calculator within seconds. However, in real life, the solutions of some problems may arise as series which do not represent any known standard functions. In those solutions, we must resort to such methods. Moreover, all iterative processes are, really speaking, meant to be carried out on computers. However, we must *understand* them so as to be able to program them for computers.

15 8·2. Evaluation of sin x

For all real values of x,

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots\ldots$$

As in case of e^x, the terms keep on decreasing and by taking the sum of a suitable number of terms, the value of $\sin x$ can be obtained to the required degree of accuracy. Notice that the nth term T_n can be written as

$$T_n = \frac{(-1)^{n+1}\, x^{2n-1}}{(2n-1)\,!}$$

To write T_{n+1}, in terms of T_n, let us compute T_{n+1}/T_n. Now

$$T_{n+1}/T_n = \frac{(-1)^{n+2}\, x^{2(n+1)-1}}{(2(n+1)-1)\,!} \times \frac{(2n-1)\,!}{(-1)^{n+1}\, x^{2n-1}},$$

$$= \frac{-x^2}{(2n+1)(2n)}$$

$$= -\frac{x^2}{2n(2n+1)}.$$

Hence $$T_{n+1} = -\frac{x^2}{2n(2n+1)} \times T_n.$$

Also, as before, denoting the sum to n terms by S_n, we get

$$S_n = S_{n-1} + T_n$$

This time let us compute sin 3 correct to three significant figures. It would be safe to compute T_n till we get four zeros after the decimal point. With $x=3$,

$$T_{n+1} = -\frac{4\cdot 5}{n(2n+1)} \times T_n$$

Our iteration starts with $n=1$. Also, the initial value $T_1 = 3$.

Evaluation of various terms of Sin 3

n	$-\dfrac{4.5}{n(2n+1)}$	T_n $(T_1=3)$	$T_{n+1}\left(=-\dfrac{4.5}{n(n+1)}\times T_n\right)$
1	−1.5000000	3.0000000	−4.5000000
2	−0.4500000	−4.5000000	2.0250000
3	−0.2142857	2.0250000	−0.4339285
4	−0.1250000	−0.4339285	0.0542410
5	−0.0818181	0.0542410	−0.0044378
6	−0.0576923	−0.0044378	0.0002560
7	−0.0428571	0.0002560	−0.0000109

Let us now tabulate S_n, T_{n+1} and S_{n+1} for various values of n in order to evaluate sin 3.

Evaluation of Sin 3

n	S_n	T_{n+1}	S_{n+1}
1	3	−4.5000000	−1.5000000
2	−1.5000000	2.0250000	0.5250000
3	0.5250000	−0.4339285	0.0910715
4	0.0910715	0.0542410	0.1453125
5	0.1453125	−0.0044378	0.1408747
6	0.1408747	0.0002560	0.1411307
7	0.1411307	−0.0000109	0.1411198

We notice that $|S_7-S_8|<10^{-4}$. Hence the required value correct to three significant figures is ·141.

Remark : We could have reduced our work somewhat had we noted that $\pi=3{\cdot}1415927$, and $\sin(\pi-\theta)=\sin\theta$. Thus $\sin 3=\sin(3{\cdot}1415927-0{\cdot}1415927)=\sin({\cdot}1415927)$. Had we taken $x=0{\cdot}1415927$, $x^2/2$ would have been ·0100242. With the value replacing 4·5 in the first column of the table for evaluating various terms of sin 3, the sought out zeros after the decimal point would have come much earlier. Such tactics which reduce the computing labour must be used whenever possible.

15·8·3. Evaluation of cos x

The power series expansion of cos x is

$$\cos x = 1 - \frac{x^2}{2\,!} + \frac{x^4}{4\,!} - \frac{x^6}{6\,!} + \ldots\ldots,$$

and it is valid for all real values of x. The general term T_n is given by

$$T_n = (-1)^n \frac{x^{2n}}{(2n)\,!}$$

Here again, we shall regard the constant term 1 as the zeroth term. Thus

$$T_0 = 1,$$

$$T_1 = -\frac{x^2}{2\,!},$$

$$T_2 = \frac{x^4}{4\,!}$$

and so on. The recurrence relation connecting T_n and T_{n+1} is

$$T_{n+1} = \frac{-x^2}{(2n+1)(2n+2)},$$

$$T_n = -\frac{x^2}{2(n+1)(2n+1)} T_n.$$

Let us evaluate cos 3. Notice that cos 3 = cos (π − 0·1415927) = − cos (0·1415927), We shall evaluate cos (0·1415927) first. When

$$x = 0{\cdot}1415927,$$

$$T_{n+1} = \frac{-0{\cdot}0100242}{(n+1)(2n+1)} T_n.$$

You can organize your work in two stages as before and check up your calculations from the following table :

Evaluation of cos (0·1415927)

n	$\dfrac{-.0100242}{(n+1)(2n+1)}$	T_n $(T_0 = 1)$	S_n $(= S_{n-1} + T_n)$ $(S_0 = 1)$
0	−0·0100242	1	1
1	−0·0016707	−0·0100242	(1·0 + (−0·0100242) =) 0·9899758
2	−0·0006682	0·0000167	0·9899925

There is no need to carry out further calculations because it is clear from the first two entries of the last row that T_3 would have

seven zeros following the decimal point, and hence would contribute nothing to the sum. The required sum is, therefore, $-0{\cdot}9899$, turncated to four digits. Rounded off to four significant digits, the value is $-0{\cdot}9900$. (Recall that $\cos 3=-\cos(0{\cdot}1415927)$.)

Remark. Compare the computational work involved here with that in evaluating sin 3. A little understanding may bring in great rewards, By using the relation $\cos(\pi-\theta)=-\cos\theta$, we sort of *well-conditioned* our input, in the sense that the sum soon *settled down* to a certain value. Using the jargon, we could say the *convergence was rapid.*

EXERCISE 15 (*i*)

Obtain all answers rounded off to three significant digits, unless stated otherwise.

1. Evaluate each of the following using iterative methods :
   (*a*) e^2. (*b*) $e^{2\cdot 5}$.
   (*c*) $e^{3\cdot 2}$. (*d*) $e^{1\cdot 4}$.
2. Use iterative methods to evaluate :
   (*a*) Sin 1. (*b*) Cos 2.
   (*c*) Sin 2·1. (*d*) Cos ·34.
   (*e*) Sin ·1415927.
3. Evaluate, making uss of trigonometric identities to reduce the arguments to a number less than $\frac{\pi}{2}$:
   (*a*) Sin 15. (*b*) Sin 100.
   (*c*) Cos 30. (*d*) Cos 78.
4. Using the series $ln\,(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots$ for $-1<x\leqslant 1$, evaluate
   (*a*) ln 1·2. (*b*) ln 1·03.
   (*c*) ln ·5. (*d*) ln 2.
5. Can you evaluate ln 3 by the series given in question 4 above ? Use $ln\left(\frac{1+x}{1-x}\right)=2\left(x+\frac{x^3}{3}+\frac{x^5}{5}+\cdots\cdots\right)$ to evaluate $ln 3$.

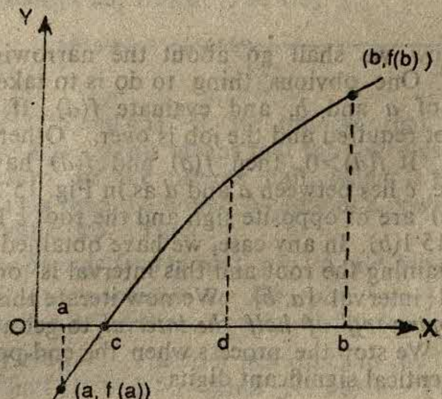
15·9. SOLUTION OF EQUATIONS

The moment any one says 'equation', what picture comes to your mind ? Is it $a_0+a_1x+\cdots\cdots+a_nx^n=0$? If yes, a word of caution is needed. In mathematics, half the time we are solving equations, but most of the time, these are not polynomial equations mentioned above. You have already come across equations of the type

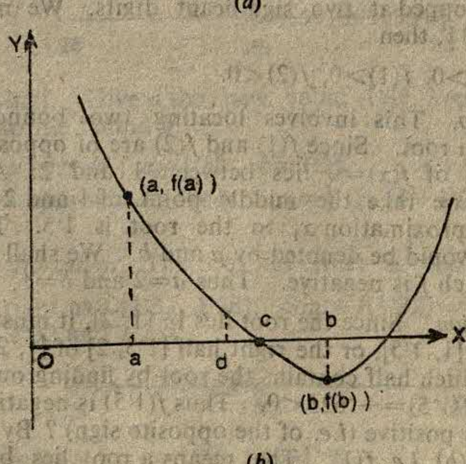
$\sin x=0$, $\tan x+2\cos x=0$, and certainly you have solved lots of differential equations. As you go to higher classes, you would discover a rich variety of equations. For the moment, we shall concentrate on equations of the type $f(x)=0$, where $f(x)$ might be a polynomial or a combination of polynomial, trigonometric and exponential functions. We shall mention some popular iterative methods to find *approximate solutions*.

15·9·1. Method of Bisection (Binary Chop)

This method is based on the *Intermediate Value Theorem* which you read in Chapter 4. Suppose f is a continuous function. Let a and b be points in the domain of f such that $f(a)$ and $f(b)$ are of opposite signs. Suppose $f(a)$ is negative and $f(b)$ is positive so that $f(a)<0<f(b)$. Since f must assume every value between $f(a)$ and



(*a*)



(*b*)

Fig. 15·1

$f(b)$ by the above theorem, and since 0 is a value between $f(a)$ and $f(b)$, $f(c)$ must be equal to 0 at some point c between a and b. This situation is shown in Fig. 15·1(a).

The other case $f(a)>0>f(b)$ is shown in Fig. 15·1(b). Once we have got a and b such that $f(a)$ and $f(b)$ are of opposite sign, we know that a root lies between a and b. Now we shall use iteration to get better and better values of a and b in the sense that they are closer and closer to c. Ultimately we shall trap the root c between two such values a and b which agree in a given number, say four, of their most significant digits. For example, if $a=2{\cdot}3417$ and $b=2{\cdot}3419$, then we have $2{\cdot}3417<c<2{\cdot}3419$. What does it mean ? It means that c is a number like 2·341......Hence chopped off at the fourth significant digit, $c=2{\cdot}341$, and this is an approximate solution.

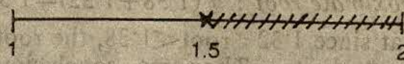
Let us see how we shall go about the narrowing business mentioned above. One obvious thing to do is to take the middle point $d=(a+b)/2$ of a and b, and evaluate $f(d)$. If $f(d)=0$ by chance, d is the root required and the job is over. Otherwise, either $f(d)>0$ or $f(d)<0$. If $f(d)>0$, then $f(a)$ and $f(d)$ have opposite signs, and therefore, c lies between a and d as in Fig. 15·1(a). Otherwise $f(d)$ and $f(b)$ are of opposite sign and the root c lies between d and b as in Fig. 15·1(b). In any case, we have obtained an interval $[a, d]$ or $[d, b]$ containing the root and this interval is only half as big as the original interval $[a. b]$. We now iterate this process of *bisecting and chopping off of half the interval* to get a yet smaller interval and so on. We stop the process when the end-points have a given number of identical significant digits.

Let us illustrate the process on the equation $x^3-10x+11=0$ to find a root chopped at two significant digits. We note that if $f(x)=x^3-10x+11$, then

$$f(0)>0,\ f(1)>0,\ f(2)<0.$$

Initial Step. This involves locating two bounds between which there lies a root. Since $f(1)$ and $f(2)$ are of opposite sign, we know that a root of $f(x)=0$ lies between 1 and 2. As a crude approximation, we take the middle point of 1 and 2 as the root. Thus our first approximation x_1 to the root is 1·5. The bounds located above would be denoted by a and b. We shall denote by a the bound at which f is negative. Thus $a=2$ and $b=1$.

First Iteration. Since the root lies in [1, 2], it must lie either in the left half [1, 1·5], or the right half [1·5, 2] of [1, 2]. We can easily discover which half contains the root by finding out the sign of $f(1{\cdot}5)$. Now $f(1{\cdot}5)=-{\cdot}625<0$. Thus $f(1{\cdot}5)$ is negative. Which of $f(1)$ and $f(2)$ is positive (*i.e.* of the opposite sign) ? By our choice of a and b, it is $f(b)$, *i.e.* $f(1)$. That means a root lies between 1·5 and 1. Hence we chop off the useless right half interval [1·5, 2] and



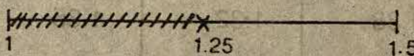
work with the new interval $[1, 1\cdot5]$ which contains the root. Since $f(1)>0$, $b=1$ and since $f(1\cdot5)<0$, $a=1\cdot5$. Thus the bound a gets the new value $1\cdot5$ and the other bound retains the old value 1. We summarize our data below :

(*a*) *new bounds for the root* : $a=1\cdot5$ and $b=1\cdot0$,

(*b*) *new approximation to the root.* $x_2=\frac{1}{2}(1\cdot0+1\cdot5)=1\cdot25$.

Notice that $1\cdot0<\text{root}<1\cdot5$ implies that the most significant digit of the root has been found. What is it ?

Second Iteration. The root lies either in the left half $[1\cdot0, 1\cdot25]$ or the right half $[1\cdot25, 1\cdot5]$ of $[1\cdot0, 1\cdot5]$.

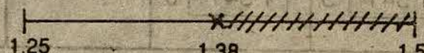


Since $f(1\cdot25)=0\cdot453>0$, we observe that $f(1\cdot25)$ and $f(1\cdot5)$ have opposite sign. This means that the root lies in $[1\cdot25, 1\cdot5]$, and we again chop off the useless half, which is the left half interval this time. This time the bound b assumes a new value $1\cdot25$ and a has the old value $1\cdot5$. Thus

(*a*) *new bounds for the root.* $a=1\cdot5$ and $b=1\cdot25$,

(*b*) *new approximation to the root.* $x_3=\frac{1}{2}(1\cdot5+1\cdot25)=1\cdot375$ $=1\cdot38$, (rounded off to three significant figures).

Third Iteration. The root lies either in the left half $[1\cdot25, 1\cdot38]$ or the right half $[1\cdot38, 1\cdot5]$. Since $f(1\cdot38)=-\cdot172<0$, $f(1\cdot38)$ and $f(1\cdot25)$ have opposite sign. Thus the root lies in the left half. Chop

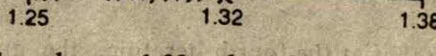


off the right half. Give a the new value $1\cdot38$ because $f(1\cdot38)<0$. Do nothing to the bound b.

(*a*) *New bounds* : $a=1\cdot38$, $b=1\cdot25$,

(*b*) *New approximate root* : $x_4=\frac{1}{2}(1\cdot38+1\cdot25)=1\cdot315=1\cdot32$ say.

Fourth Iteration. The root lies either in the left half $[1\cdot25, 1\cdot32]$ or the right half $[1\cdot32, 1\cdot38]$. Since $f(1\cdot32)=0\cdot1>0$, the root lies between $1\cdot32$ and $1\cdot38$.



(*a*) *New bounds* : $a=1\cdot38$, $b=1\cdot32$.

(*b*) *New approximate root* : $\frac{1}{2}(1\cdot38+1\cdot32)=1\cdot35$.

We notice that since $1\cdot32<\text{root}<1\cdot38$, the root is $1\cdot3$ to the two most significant digits. The same conclusion is reached by comparing the two latest approximations x_4 and x_5 which agree in the two most significant digits. Hence chopped off at the second digit, the root is $1\cdot3$.

Let us now examine how many iterations are required to locate the third and the fourth significant digit. Skipping the details, we list below the relevant computations :

Iteration	a	b	$\frac{1}{2}(a+b)$	$f(\frac{1}{2}(a+b))$
5 th	1.38	1.32	1.35	< 0
6 th	1.35	1.32	1.335	> 0
7 th	1.35	1.335	1.3425	< 0
8 th	1.3425	1.335	1.3387	> 0
9 th	1.3425	1.3387	1.3406	> 0
10 th	$\boxed{1.34}$ 25	$\boxed{1.34}$ 06	1.3415	< 0
11 th	1.3415	1.3406	1.3410	> 0 (= .0014)
12th	$\boxed{1.341}$ 5	$\boxed{1.341}$ 0		

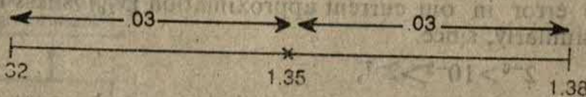
It follows that 9 iterations are needed to locate the third significant digit and 11 for the fourth. The root after the 11th iteration is $1\cdot341$ and $f(1\cdot341)=\cdot0014$.

Remarks. 1. Since we want the root chopped at the second digit, we do not need the value beyond three significant digits. That is why we rounded $1\cdot375$ to $1\cdot38$ at the second iteration. Had we wanted greater accuracy, we would keep the values intact to more significant digits.

2. At every stage, it is not really necessary to compute the actual value $f(x_n)$. All we are interested in is knowing the sign of $f(x_n)$.

3. Notice that at the fourth iteration, the approximate root is $1\cdot35$ and it is the middle point of $[1\cdot32, 1\cdot38]$. We know for

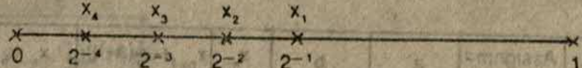
certain that the actual root lies in this interval. Hence the last approximation 1·35 has a maximum absolute error ·03 (equal to



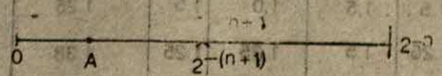
half the length of the interval [1·32, 1·38]. Since one of the end-points is always the previous approximation, the maximum absolute error is equal to the magnitude of the difference between the last two approximations. Hence a good indicator to stop the iterations is the *difference between two successive approximations*. Whenever this becomes less than a pre-assigned number, we can stop iterating, and take the last approximation as the required value.

4. Each iteration halves the maximum absolute error. Equivalently, each iteration increases the precision by a facto of 2, or one binary digit. Hence this method is known as the **binary chop.** This method is named after Bolzano as well.

5. Though *sure*, this method is *slow*. In general, it takes more than ten iterations to find the root to any reasonable accuracy. In fact, if we start with an interval of length 1, it is only after ten iterations that we find the error in our approximation to be less than 10^{-3}. To see this, let us suppose that the root is known to lie between 0 and 1 and that at each step, the right half of the interval is chopped off. Our first approximation to the root is $x_1=\frac{1}{2}(0+1)=2^{-1}$. After the first iteration the root is known to lie somewhere in the interval $I_1=[0, 2^{-1}]$ of length 2^{-1} and our second approximation to the root is $x_2=2^{-2}$.



After the second iteration, the root is known to lie somewhere in the interval $I_2=[0, 2^{-2}]$ of length 2^{-2} and our third approximation $x_3=2^{-3}$. Thus in general, after the nth iteration, the root is known to lie somewhere in the interval $I_n=[0, 2^{-n}]$ of length 2^{-n}, say at A, and our $(n+1)$th approximation is $x_{n+1}=2^{-(n+1)}$. Clearly the maximum error in x_{n+1} is $2^{-(n+1)}$.



Thus it is only after n iterations that we can say with confidence that the error in the current approximation x_{n+1} is less than $2^{-(n+1)}$. Now

$$2^9(=512)<1000<2^{10}(=1024),$$

So that, $\quad 2^{-9} > 10^{-3} > 2^{-10} = 2^{-(9+1)}$.

Thus it is only after 9 iterations in general, that we can say that the error in our current approximation (x_{10}) is less than 10^{-3} or ·001 similarly, since

$$2^{-6} > 10^{-2} > 2^{-7},$$

it is only after 6 iterations in general that we can be sure of error being less than 10^{-2} or ·01. How many iterations atleast would be required if we want to be sure that the error is less than 10^{-1} or ·1 ? Three !

6. It makes life easier if we evaluate expressions like $a_0x^3 + a_1x^2 + a_2x + a_3$ as a nested multiplication

$$x\{x(a_0x + a_1) + a_2\} + a_3,$$

starting the computations from the inner-most brackets. Also, it is more efficient to evaluate

$$\tfrac{1}{2}(a+b) \text{ as } a + \tfrac{1}{2}(b-a), \text{ or } b + \tfrac{1}{2}(a-b).$$

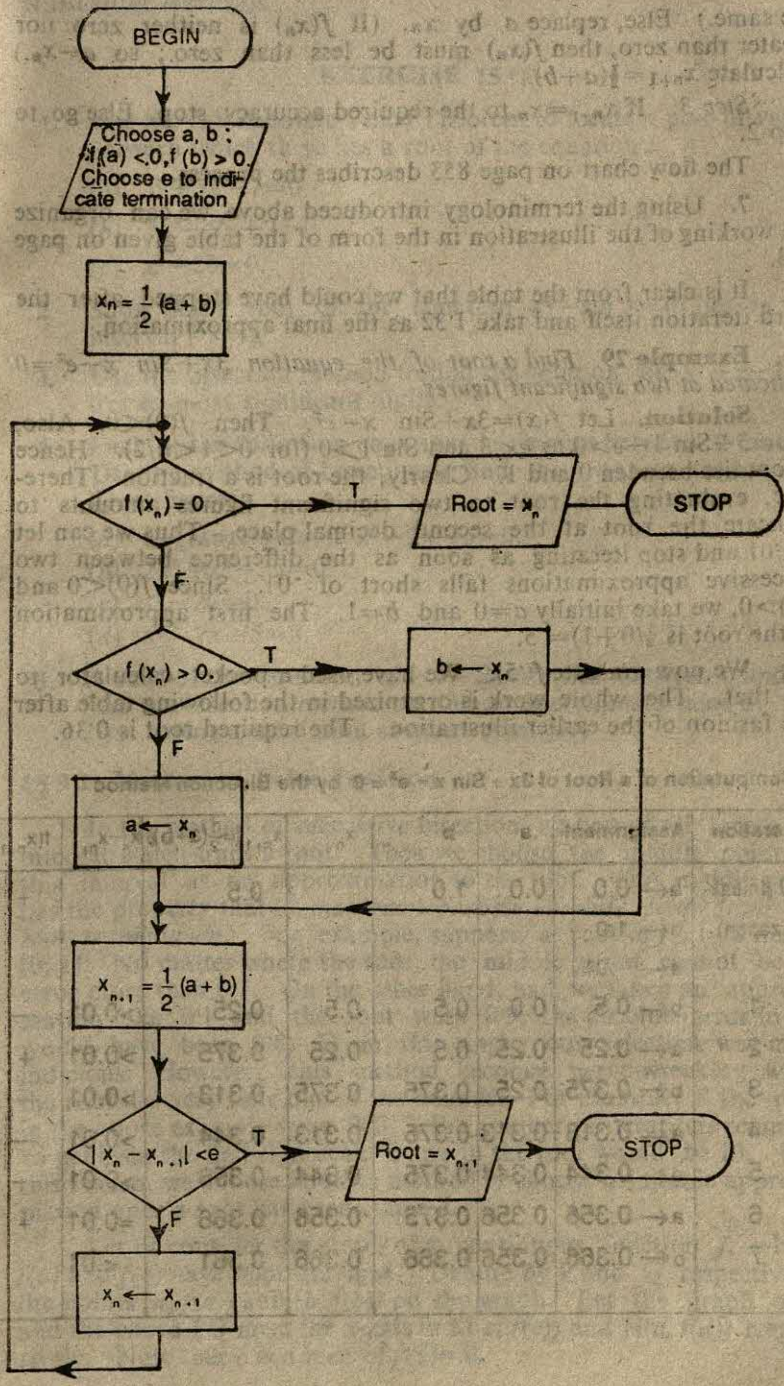
7. We can now formalize what we have done above as follows :

Step 1. Find values a and b such that $f(a) < 0$ and $f(b) > 0$. Take the first approximation to the root as $\frac{1}{2}(a+b)$.

Step 2. (nth iteration) If $f(x_n) = 0$, stop. x_n is the root we are trying to trap. Else, test the sign of $f(x_n)$. If $f(x_n) > 0$ replace b by x_n. (Recall that the bound at which f is positive is labelled b, and that x_n replaces one of the bounds, the other bound remaining

Computation of a Root of $x^3 - 10x + 11 = 0$ by the Bisection Method

Iteration	Assignment	a	b	x_n	$x_{n+1}(=\frac{1}{2}(a+b))$	$x_n - x_{n+1}$	$f(x_{n+1})$
0 (Initialization)	$a \leftarrow 2.0$ $b \leftarrow 1.0$ $e \leftarrow 0.1$	2.0	1.0		1.5		−
1	$a \leftarrow 1.5$	1.5	1.0	1.5	1.25	.25 (> .1)	+
2	$b \leftarrow 1.25$	1.5	1.25	1.25	1.38	.13 (> .1)	—
3	$a \leftarrow 1.38$	1.38	1.25	1.38	1.32	.06 (< .1)	+
4	$b \leftarrow 1.32$	1.38	1.32	1.32	1.35		



Bisection method to solve f(x)=0.

the same.) Else, replace a by x_n. (If $f(x_n)$ is neither zero nor greater than zero, then $f(x_n)$ must be less than zero; so $a \leftarrow x_n$.) Calculate $x_{n+1}=\frac{1}{2}(a+b)$.

Step 3. If $x_{n+1}=x_n$ to the required accuracy, stop. Else go to step 2.

The flow chart on page 853 describes the process.

7. Using the terminology introduced above, we can organize the working of the illustration in the form of the table given on page 853.

It is clear from the table that we could have stopped after the third iteration itself and take 1·32 as the final approximation.

Example 29. *Find a root of the equation $3x+\text{Sin } x-e^x=0$ truncated at two significant figures.*

Solution. Let $f(x)=3x+\text{Sin } x-e^x$. Then $f(0)<0$. Also, $f(1)=3+\text{Sin } 1-e>0$ as $e<3$ and Sin $1>0$ (for $0<1<\pi/2$). Hence a root lies between 0 and 1. Clearly, the root is a fraction. Therefore, evaluating the root to two significant figures amounts to truncate the root at the second decimal place. Thus we can let $e=\cdot01$ and stop iterating as soon as the difference between two successive approximations falls short of ·01. Since $f(0)<0$ and $f(1)>0$, we take initially $a=0$ and $b=1$. The first approximation to the root is $\frac{1}{2}(0+1)=\cdot5$.

We now evaluate $f(\cdot5)$. We have used a pocket calculator to do that. The whole work is organized in the following table after the fashion of the earlier illustration. The required root is 0·36.

Computation of a Root of $3x + \text{Sin } x - e^x = 0$ by the Bisection Method

Iteration	Assignment	a	b	x_n	$x_{n+1}\ (=\frac{1}{2}(a+b))$	x_n-x_{n+1}	$f(x_{n+1})$
0 (Initialization)	a← 0.0 b← 1.0 e← .01	0.0	1.0		0.5		+
1	b← 0.5	0.0	0.5	0.5	0.25	>0.01	—
2	a← 0.25	0.25	0.5	0.25	0.375	>0.01	+
3	b← 0.375	0.25	0.375	0.375	0.313	>0.01	—
4	a← 0.313	0.313	0.375	0.313	0.344	>0.01	—
5	a← 0.344	0.344	0.375	0.344	0.356	>0.01	—
6	a← 0.356	0.356	0.375	0.356	0.366	=0.01	+
7	b← 0.366	0.356	0.366	0.366	0.361	<.01	

EXERCISE 15 (j)

1. Use the Intermediate Value Theorem to locate a pair of values between which there lies a root of the equation :

   (a) $x^3+3x-7=0$

   (b) $3x^3-9x+2=0$

   (c) $x^3-20=0$.

2. Show that a root of the equation $x^2-\sin x=0$ lies in the interval [·5, 1].

3. Use the bisection method to find the solution of $x^3-x-1=0$ to two most significant digits.

4. Each of the following equations has a root between 2 and 3. Use the method of bisection to find this root to three significant figures :

   (a) Lagrange's Cubic $x^3-2x-5=0$.

   (b) $2x^3+x^2-33=0$.

   (c) $x^3-26=0$.

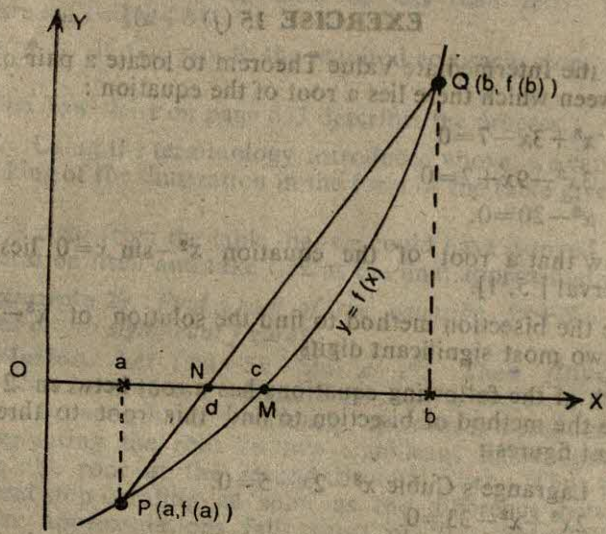
   (d) $x^3-7x+5=0$.

5. A root of the equation $x^2-x-·24=0$ lies between 0 and ·5. Use the bisection method to find an approximate value of this root having a maximum absolute error ·01.

15·9·2 Method of False Position

In the method of successive bisection, we first of all locate an interval which traps a root. Then we choose the middle point of this interval as an approximation to the root. This middle point has the property that its *maximum possible absolute error from the root is minimum.* For example, suppose a root of $f(x)=0$ lies in [0, 1]. No matter where the root, the middle point cannot be in error more than 0·5. On the other hand, had we taken an approximation like 0·1 and the root were 0·9, the absolute error in 0·1 would have been 0·8. From this angle, our selection was most judicious. However, this method becomes nerve-wrecking when the root lies very near one of the bounds. For example, if the root in the above example were ·00015, the number of iterations required to get even the first significant figure in the root would be 13. For this reason, we choose another method to choose a suitable approximation once a root has been trapped.

Let us consider the graph of a continuous function f, where $f(a)$ and $f(b)$ have opposite signs. Denote by P and Q respectively, the points $(a, f(a))$ and $(b, f(b))$ on the graph. Let the graph of f and the chord PQ meet the x-axis in M$(c, f(c))$ and N$(d, f(d))$ respectively. Note that c is a root of $f(x)=0$.



We assume that between a and b, the graph is approximated by the chord PQ so that d may be taken as an approximate value of c. The point N can be easily located as the intersection of the chord PQ with the x-axis.

Having obtained d, we start iterating. As in the bisection method, we narrow down the interval containing the root with the help of d by taking d as one of the bounds. As there, we shall call that bound at which f is negative, a, and the other one b. Having obtained a new pair of bounds, we again find the point of the intersection of the x-axis with chord joining the new points on the graph corresponding to the new bounds. The process is continued till a root with the desired accuracy is found. Let us illustrate the process by solving the same transcendental equation $3x+\sin x-e^x=0$ which we took in Example 29.

Here $f(x)=3x+\sin x-e^x$, $f(0)=-1<0$, $f(1)=1{\cdot}12>0$. Adopting the previous practice of taking a as the bound at which f is negative, we have $a=0$, $b=1$. The equation of the line joining $(a, f(a))$ and $(b, f(b))$, *i.e.*, $(0, -1)$ and $(1, 1{\cdot}12)$ is

$$y+1=\frac{1{\cdot}12+1}{1-0}(x-0),$$

or

$$y+1=2{\cdot}12x.$$

This meets the x-axis where $x=1/2{\cdot}12={\cdot}47$ (rounding off to two decimal places). Thus our first approximation to the root is ${\cdot}47$.

Now we start our first iteration. $f({\cdot}47)=0{\cdot}26<0$. Hence the bound b is replaced by the approximation $0{\cdot}47$; a has the previous value 0. Thus we now know that the root lies in $[0, {\cdot}47]$. The

equation of the chord through $(a, f(a))$ and $(b, f(b))$ is now the equation of the line joining the points $(0, -1)$, and $(0{\cdot}47, 0{\cdot}26)$, and is given by

$$y+1=\frac{0{\cdot}26+1}{0{\cdot}47-0}(x-0),$$

or
$$y+1=\frac{1{\cdot}26}{0{\cdot}47}x.$$

This meets the x-axis where $x=\frac{47}{126}=0{\cdot}373$ (taking one more decimal place). Thus the second approximation is 0.373.

Now $f(0{\cdot}373)=0{\cdot}031<0$. Hence we replace b by the latest approximation $0{\cdot}373$. Thus the root lies in $[0, 0{\cdot}373]$. The equation of the current chord passing through $(0, -1)$ and $(0{\cdot}373, 0{\cdot}031)$ is

$$y+1=\frac{0{\cdot}031+1}{0{\cdot}373}x,$$

and meets the x-axis where

$$x=\frac{0{\cdot}373}{1{\cdot}031}=0{\cdot}361.$$

Hence the third approximation is $0{\cdot}361$. (The first significant figure in the root has been found, the two latest approximations being $0{\cdot}373$ and $0{\cdot}361$.) $f({\cdot}361)=0{\cdot}001<0$. Hence b is assigned the value $0{\cdot}361$.

The equation of the new chord (through $(0, -1)$ and $(0{\cdot}361, 0{\cdot}001)$ is

$$y+1=\frac{0{\cdot}001+1}{0{\cdot}361}x,$$

and it meets the x-axis where

$$x=\frac{0{\cdot}361}{1{\cdot}001}=0{\cdot}360.$$

Hence $0{\cdot}360$ is the fourth approximation. Since the previous approximation was $0{\cdot}361$, we now know that the root, chopped at the second decimal place is $0{\cdot}36$.

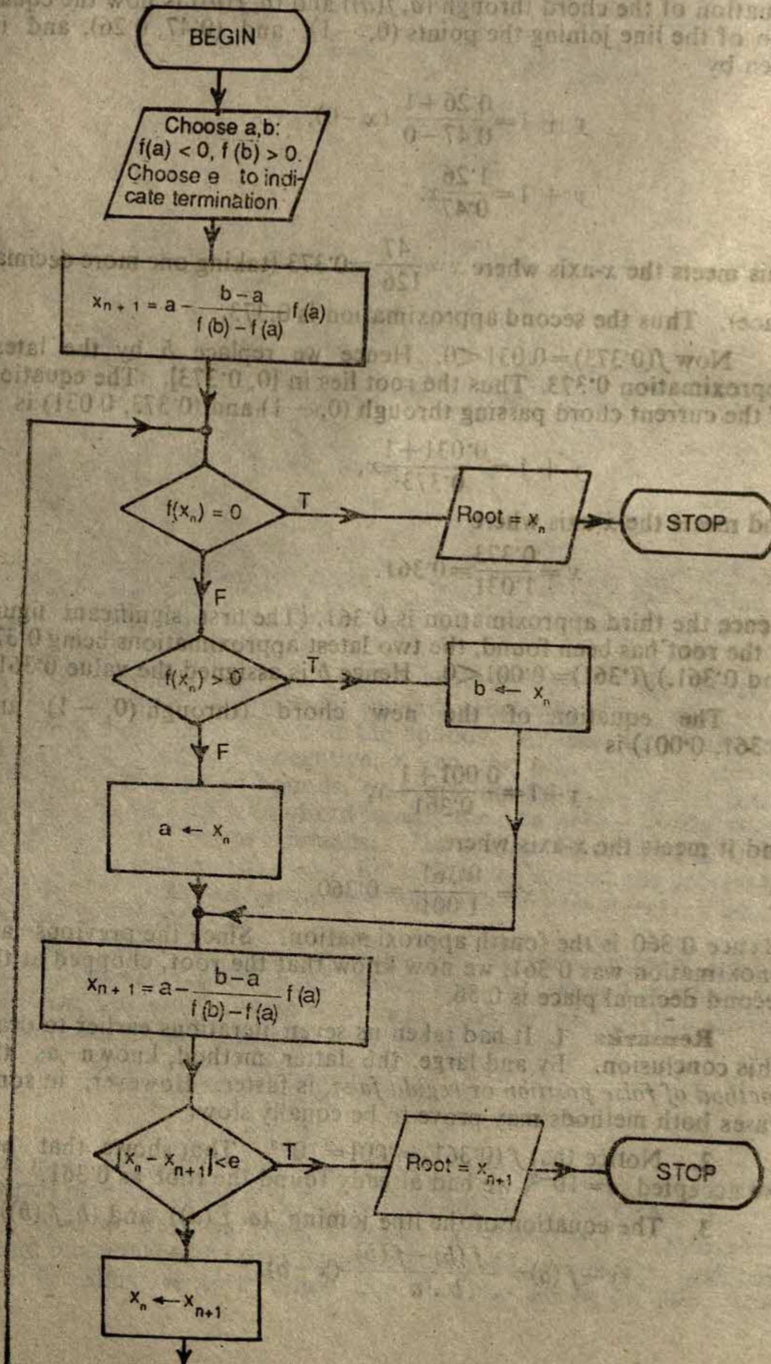
Remarks. 1. It had taken us seven iterations earlier to draw this conclusion. By and large, the latter method, known as the *method of false position* or *regula falsi*, is faster. However, in some cases both methods may prove to be equally slow.

2. Notice that $f(0{\cdot}361)={\cdot}001=10^{-3}$. That shows that had we accepted $e=10^{-2}$, we had already found the root as $0{\cdot}361$.

3. The equation of the line joining $(a, f(a))$ and $(b, f(b))$ is

$$y-f(a)=\frac{f(b)-f(b)}{b-a}(x-a).$$

Finding a root of f(x)=0 by the method of false position.



It meets the x-axis where

$$x=a-\frac{b-a}{f(b)-f(a)}f(a), \qquad \ldots(1)$$

$$=\frac{af(b)-bf(a)}{f(b)-f(a)} \qquad \ldots(2)$$

Instead of finding the various approximations by first finding the equation of the chord, and then the point of intersection with the x-axis, we can use the above formula directly. As far as computer is concerned, (1) is preferable as it involves only one multiplication.

Let us now formulate the method of false position :

Step 1. (Initialization) : Find two values a and b such that $f(a)<0$ and $f(b)>0$. The first approximation x_1 to the root is

$$a-\frac{b-a}{f(b)-f(a)}f(a)$$

Step 2. (*n*th Iteration) : Evaluate $f(x_n)$. If $f(x_n)=0$, x_n is the required root. If $f(x_n)>0$, replace b by x_n. Else, replace a by x_n. Evaluate the next approximation x_{n+1} by the formula

$$x_{n+1}=a-\frac{b-a}{f(b)-f(a)}f(a).$$

Step 3. If $x_n=x_{n+1}$ to the required accuracy, stop. Else, go to step 2.

The flow chart on page 858 describes the process.

Example 30. *Find the root of $x^3-2x-5=0$ which lies between 2 and 3, correct to three places of decimal.*

Solution. Since we wish to find the root correct to three decimal places, we shall calculate the root to four decimal places. Now $f(2)=-1, f(3)=16$. Thus $a=2$ and $b=3$.

The first approximation x_1 to the root is

$$x_1=a-\frac{b-a}{f(b)-f(a)}f(a),$$

$$=2-\frac{3-2}{16-(-1)}(-1),$$

$$=2+\frac{1}{17}=2{\cdot}06$$

(rounded off to 2 decimal places).

First Iteration. $f(2{\cdot}06)=-{\cdot}38<0$. We replace a by $2{\cdot}06$. b has the old value 3. The second approximation x_2 is given by

$$x_2 = 2{\cdot}06 - \frac{3-2{\cdot}06}{16-(-{\cdot}38)}(-{\cdot}38),$$
$$= 2{\cdot}06 + \frac{{\cdot}94 \times {\cdot}38}{16{\cdot}38}$$
$$= 2{\cdot}082 \text{ (rounding off to 3 decimal places).}$$

We have taken three decimal places here because the root has started emerging as 2·0... Clearly $|x_1 - x_2| > e$.

Second Iteration. $f(2{\cdot}082) = -0{\cdot}1391 < 0$. So we replace a by 2·082 ; b is 3 still. The third approximation x_3 is given by

$$x_3 = 2{\cdot}082 - \frac{3-2{\cdot}082}{16+0{\cdot}1391}(-0{\cdot}1391),$$
$$= 2{\cdot}082 + \frac{0{\cdot}918 \times 0{\cdot}1391}{16{\cdot}1391} = 2{\cdot}090.$$

Third Iteration. $f(2{\cdot}09) = -0{\cdot}0506 < 0$. Hence we replace a by 2·09. The fourth approximation x_4 is given by

$$x_4 = 2{\cdot}09 - \frac{3-2{\cdot}09}{16+0{\cdot}0506}(-0{\cdot}0506),$$
$$= 2{\cdot}09 + \frac{0{\cdot}91 \times 0{\cdot}0506}{86{\cdot}0506} = 2{\cdot}0928.$$

Fourth Iteration. $f(2{\cdot}0928) = -0{\cdot}01952 < 0$. Replace a by 2·0928.

$$x_5 = 2{\cdot}0928 - \frac{3-2{\cdot}0928}{16+0{\cdot}01952}(-0{\cdot}01952),$$
$$= 2{\cdot}0928 + \frac{0{\cdot}9072 \times 0{\cdot}01952}{16{\cdot}01952}$$
$$= 2{\cdot}09390.$$

Fifth Iteration. $f(2{\cdot}0939) = -0{\cdot}00726 < 0$. Hence $a \leftarrow 2{\cdot}09390$.

$$x_6 = 2{\cdot}09390 - \frac{3-2{\cdot}09390}{16+0{\cdot}00726}(-0{\cdot}00726),$$
$$= 2{\cdot}09390 + \frac{0{\cdot}9061 \times 0{\cdot}00726}{16{\cdot}00726}$$
$$= 2{\cdot}09431.$$

Sixth Iteration. $f(2{\cdot}09431) = -0{\cdot}00270 < 0$. Hence $a \leftarrow 2{\cdot}09431$.

$$x_7 = 2{\cdot}09431 + \frac{0{\cdot}90569}{16{\cdot}00270} \times 0{\cdot}00270,$$
$$= 2{\cdot}09431 + 0{\cdot}00015,$$
$$= 2{\cdot}09446.$$

Seventh Iteration. $f(2\cdot09446)=-0\cdot00103<0$.
Hence $a \leftarrow 2\cdot09446$.

$$x_8=2\cdot09446+\frac{3-2\cdot09446}{16+0\cdot00103}\times 0\cdot00103,$$
$$=2\cdot09446+\cdot00005=2\cdot09451.$$

Eighth Iteration. $f\ (2\cdot09451)=-\cdot00047<0$.
Hence $a \leftarrow 2\cdot09451$.

$$x_9=2\cdot09451+\frac{3-2\cdot09451}{16+0\cdot00047}\times 0\cdot00047,$$
$$=2\cdot09451+\cdot00002=2\cdot09453.$$

Hence the root is 2·0945 to four decimal places. Rounding off to get the value correct to three decimal places, we get 2·094 as the root. Recall that 4 being even, is not to be increased because of the 5 being replaced by zero.

Computation of a Root of $x^3-2x-5=0$ by the Method of False Position.

Itera tion	Assign-ment	a	f(a)	b	f(b)	X_{n+1} $=a-\frac{b-a}{f(b)-f(a)}f(a)$	$f(x_{n+1})$
0(Initialization)	$\left.\begin{array}{l} a \leftarrow 2 \\ b \leftarrow 3 \end{array}\right\}$	2.0	−1.0	3.0	16.0	2.06	-
1	$a \leftarrow 2.06$	2.06	−.38	3.0	16.0	2.082	-
2	$a \leftarrow 2.082$	2.082	−.1391	3.0	16.0	2.090	-
3	$a \leftarrow 2.090$	2.090	−.0506	3.0	16.0	2.0928	-
4	$a \leftarrow 2.0928$	2.0928	−.01952	3.0	16.0	2.09390	-
5	$a \leftarrow 2.09390$	2.09390	.00726	3.0	16.0	2.09431	-
6	$a \leftarrow 2.09431$	2.09431	.00270	3.0	16.0	2.09446	-
7	$a \leftarrow 2.09446$	2.09446	.00103	3.0	16.0	$2.\underbrace{0945}1$	-
8	$a \leftarrow 2.09451$	2.09451	.00047	3.0	16.0	$2.\underbrace{0945}3$	-

EXERCISE 15 (*k*)

1. Obtain a root, of each of the following equations, truncated at three decimal places by the method of false position :

   (*a*) $x^3-4x-9=0$.

   (*b*) $x^3-18=0$.

   (*c*) $x^3+x^2-3x-3=0$.

   (*d*) $x^3-x-1=0$.

2. The equation $x^3+3x-7=0$ has a root between 1·4 and 1·6. Obtain the root correct to three places of decimal by using the method of false position.

3. Find a cube root of 5 correct to one decimal place by making use of the method of false position.

4. A root of $xe^x-3=0$ lise between 1·0 and 1·1. Use the method of false position to find this root correct to two decimal places.

5. A root of $\sin xe^x=1$ lies between ·55 and ·5. Find this root truncated at the third decimal place.

15 9.3. Newton-Raphson Method

If f and f' are continuous functions on the closed interval $[a, a+h]$ and f'' exists in $]a, a+h[$, then we can easily deduce from Rolle's theorem that there exists a positive real number θ lying between 0 and 1 such that

$$f(a+h)=f(a)+hf'(a)+(h^2/2)f''(a+\theta h). \qquad \text{...(A)}$$

The above result can be used with advantage in finding out good approximations to roots of $f(x)=0$ very easily under certain circumstances. Suppose a is an approximation to a root of $f(x)=0$ which has a *small error*. Suppose that the actual root is $a+h$, so that h is a small number. Now $a+h$ being a root of $f(x)=0$, we must have $f(a+h)=0$. But the equation (A) above then implies that

$$f(a)+hf'(a)+(h^2/2)f''(a+\theta h)=0, \qquad \text{...(B)}$$

Since h is a small number, h^2 is relatively smaller as compared to h. Hence we may ignore the last term in (B) and determine to h. from the resulting equation

$$f(a)+hf'(a)=0. \qquad \text{...(C)}$$

The value $h_0=-f(a)/f'(a)$ obtained from (C) is not the actual value of h satisfying (B), but the error in h_0 is negligible, h^2 being very small. Hence if a is an appraximate root, a much better approximation to the root $a+h$ is

$$a+h_0=a-\frac{f'(a)}{f'(a)}.$$

We can now iterate the above process starting with the new approximation $a-\dfrac{f(a)}{f'(a)}$, and find a better approximation. The process can be iterated till a root to desired accuracy is obtained.

We can use a geometrical argument also to get the above process which is known as the **Newton-Raphson Method.**

In the method of false position, we approximate the graph by a chord. Instead of a chord, we shall now use a tangent to the

graph as an approximation. Let us draw the graph of $f(x)=0$. If it meets the x-axis at the point $Q(c, 0)$, then c is a root of $f(x)=0$. Suppose we are able to a locate a value a which is *not far from* c. The corresponding point the graph is $P(a, f(a))$ and the equation of the tangent at P is

$$y-f(a)=f'(a)(x-a),$$

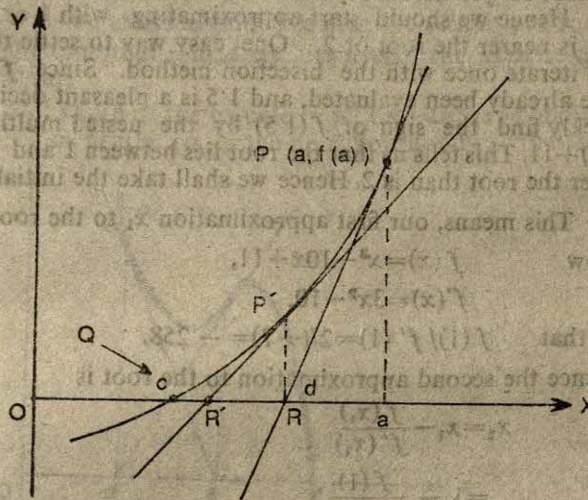


Fig. 15·2.

and it meets the x-axis at $R(d, 0)$ where

$$d=a-\frac{f(a)}{f'(a)}.$$

Since a and c are closely by, we may approximate the graph of f between P and Q by straight line segment PR. But then d is an approximation to the root c. (See Fig. 15·2). Hence as before,

$$d=a-\frac{f(a)}{f'(a)}$$

is a better approximation than a. When can now iterate starting with d, get the corresponding point P′ on the graph, approximate the graph between P′ Q by the straight line segment P′R′, R′ being the point where the tangent at P′ meets the x-axis, and so on so forth.

Newton-Raphson method, whenever applicable, converges very fast. Alas ! there are certain booby-traps, which we shall point out a little later. Let us for the moment solve a problem, met with earlier, by Newton-Raphson method.

Example 31. *Find a root of $x^3-10x+11=0$ by using Newton-Raphson method.*

Solution. As before, we decide easily that a root lies between 1 and 2. We pause here to consider the question :

Is the root nearer 1 or nearer 2 ? Remember that the success of Newton-Raphson method depends on the fact that our initial value a must be quite close to the root so that the error h in a may be small. Hence we should start approximating with 1 or 2 according as 1 is nearer the root or 2. One easy way to settle this question is to iterate once with the bisection method. Since $f(1)$ and $f(2)$ have already been evaluated, and 1·5 is a pleasant decimal, we may quickly find the sign of $f(1·5)$ by the nested multiplication $x(x^2-10)+11$. This tells us that the root lies between 1 and 1·5. Thus 1 is nearer the root than is 2. Hence we shall take the initial value as

1. This means, our first approximation x_1 to the root is 1.

Now $$f(x)=x^3-10x+11,$$

$$f'(x)=3x^2-10,$$

so that $$f(1)/f'(1)=2/(-7)=-·258.$$

Hence the second approximation to the root is

$$x_2=x_1-\frac{f(x_1)}{f'(x_1)}$$

$$=1-\frac{f(1)}{f'(1)},$$

$$=1-(-·285)=1·285.$$

The third approximation x_3 is given by

$$x_3=x_2-\frac{f(x_2)}{f'(x_2)},$$

$$=1·285-\frac{f(1·285)}{f'(1·285)},$$

$$=1·285-\frac{0·271}{-5·046},$$

$$=1·285+0·053=1·338.$$

So far so good. After the fourth iteration in the previous solution, we had discovered that the root lies in [1·32, 1·38]. The fourth approximation x_4 is given by

$$x_4=x_3-\frac{f(x_3)}{f'(x_3)},$$

$$=1·338-\frac{0·015}{-4·629},$$

$$=1·338+0·003=1·341.$$

How easy life is with Newton ! What had required eleven iterations earlier, has taken us only three here. In fact *when Newton's method is good, it is very very good, but when it is bad it is horrid.* You should never apply Newton's method (except once to convince yourself of the horrors it may cause !) under the following circumstances :

(*a*) When the derived function is complicated to compute.

(*b*) When you are trying to locate a multiple root c.

(*c*) When an extreme value lies near the sought out root c.

(*d*) When another root d say, lies nearby the root c you are seeking.

The root cause of the trouble in all the above conditions is that $f'(x) = 0$ or nearly so, in the neighbourhood of the approxi-

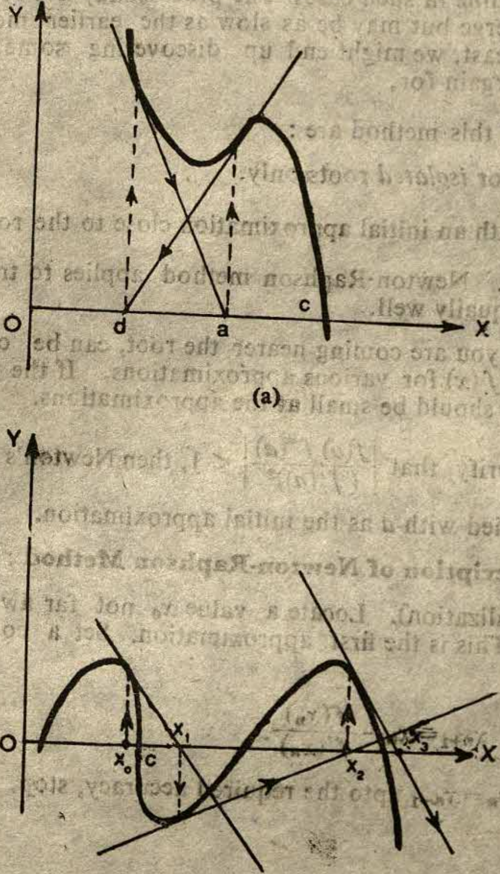


Fig. 15.3.

mations we are taking. In case of (a), $f'(c)=0$ and since f is continuous, in some neighbourhood of c, $1/f'(x)$ is too large so that the next approximation $x-(f(x)/f'(x))$ shoots off far away from c. At an extreme value, $f'(x)=0$. Thus we encounter the same type of difficulty in case of (b) too. Moreover, we may end up end up in endless cycling from one approximation to the next as in Fig. 15·3(a). Instead of coming nearer and nearer the root, we might as well be marching farther and farther away from the root as in Fig. 15·3 (b). The successive approximations are shown by x_0, x_1, x_2......... In case of (c) above, since $f(c)=f(d)=0$, by Rolle's theorem $f'(x)=0$ somewhere between c and d. Again we encounter our old away.

Another kind of foe awaits us when we start with an initial value far off from the root. It may be quite unjustified to treat the graph as a straight line in such cases. The process may not converge at all. It may converge but may be as slow as the earlier methods. Last, but not the least, we might end up discovering some other root we did not bargain for,

The DO'S of this method are :

(*i*) Use it for *isolated* roots only.

(*ii*) Start with an initial approximation close to the root.

Remarks. 1. Newton-Raphson method applies to transcendental equations equally well.

2. Whether you are coming nearer the root, can be checked from the values of $f(x)$ for various approximations. If the process is converging, $f(x)$ should be small at the approximations.

3. If you verify that $\left|\dfrac{f(a)\,f''(a)}{(f'(a))^2}\right| < 1$, then Newton's method may be safely applied with a as the initial approximation.

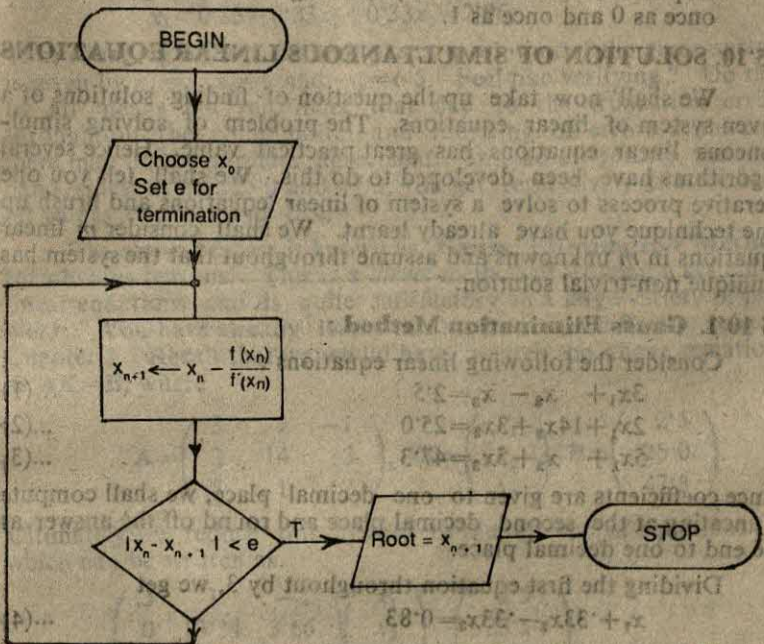
The Formal Description of Newton-Raphson Method :

Step 1 (Initialization). Locate a value x_0 not far away from an isolated root. This is the first approximation. Set a constant e for termination.

Step 2. $x_{n+1}=x_n-\dfrac{f(x_n)}{f'(x_n)}$.

Step 3. If $x_n=x_{n-1}$ upto the required accuracy, stop. Otherwise repeat step 2.

Flow-Chart Describing Newton-Raphson Method



EXERCISE 15 (*i*)

1. In each case find a real root correct to three places of decimal:
   (*a*) $x^4-x-9=0$. (*a*) $x^3-216=0$.
   (*c*) $xe^x=0$. (*d*) $x^3+x^2+x-100=0$.
   (*e*) $x^3-3x-3=0$. (*f*) $x^3-x-1=0$.

2. Show that Newton-Raphson's iteration formula $x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}$ leads to the iteration formula $x_{n+1}=\frac{1}{2}(x_n+a/x_n)$ to find the positive square root of a positive number a.
   [**Hint.** Let $f(x)=x^2-a$. Then $\sqrt{a}$ is a root of $f(x)=0$.]

3. Apply the iteration formula in the previous exercise to determine $\sqrt{3}$, $\sqrt{5}$ and $\sqrt{7}$ to four places of decimal.

4. The equation $x \sin x+\cos x=0$ has a root near π. Use Newton Raphson's method to find this root truncated at four decimal places.

5. A root of $\sin x+x=1$ lies in the neighbourhood of 0·5. Find the root correct to three decimal places.

6. A root of $2e^{-x}-\sin x=0$ lies between 1 and 2. Use Newton-Raphson's algorithm to find it, taking the initial approximation once as 0 and once as 1.

15·10. SOLUTION OF SIMULTANEOUS LINEAR EQUATIONS

We shall now take up the question of finding solutions of a given system of linear equations. The problem of solving simultaneous linear equations has great practical value. Hence several algorithms have been developed to do this. We shall tell you one iterative process to solve a system of linear equations and brush up one technique you have already learnt. We shall consider m linear equations in m unknowns and assume throughout that the system has a unique non-trivial solution.

15·10·1. Gauss Elimination Method :

Consider the following linear equations :

$$3x_1+\ x_2-\ x_3=2{\cdot}5 \qquad \ldots(1)$$

$$2x_1+14x_2+3x_3=25{\cdot}0 \qquad \ldots(2)$$

$$5x_1+\ x_2+5x_3=47{\cdot}3 \qquad \ldots(3)$$

Since coefficients are given to one decimal place, we shall compute truncating at the second decimal place and round off the answer at the end to one decimal place.

Dividing the first equation throughout by 3, we get

$$x_1+{\cdot}33x_2-{\cdot}33x_3=0{\cdot}83 \qquad \ldots(4)$$

Multiplying (4) by 2 and subtracting it from (2) gives

$$13{\cdot}34x_2+3{\cdot}66x_3=23{\cdot}34 \qquad \ldots(5)$$

Multiplying (4) by 5 and subtracting from (3), we get

$$-{\cdot}65x_2+6{\cdot}65x_3=43{\cdot}15 \qquad \ldots(6)$$

The given system of equations has the same solution as that of (1), (5) and (6).

In effect, we have *eliminated* x_1 from equations (2) and (3) and have replaced them by equations (5) and (6). We can easily *eliminate* x_2 from equation (6). Dividing (5) throughout by 13·34, multiplying the resulting equation by 0·65 and adding it to (6) we get

$$6{\cdot}82x_3=44{\cdot}28 \qquad \ldots(7)$$

The given system of equations is equivalent to

$$\left.\begin{aligned} 3x_1+x_2-x_3&=2{\cdot}5\\ 13{\cdot}34x_2+3{\cdot}66x_3&=23{\cdot}34\\ 6{\cdot}82x_3&=44{\cdot}28 \end{aligned}\right\} \qquad \ldots(A)$$

From the last of these, we get $x_3=6{\cdot}49$. Substituting this value of x_3 in (6), we get

$$x_2=0{\cdot}01.$$

Substituting these value of x_2 and x_3 in equation (4), we get

$$x_1=0{\cdot}83-0{\cdot}33x_2+0{\cdot}33x_2=2{\cdot}96.$$

Thus rounding off to 1 decimal place, a solution of the given system is given by $x_1=3$, $x_2=0$ and $x_3=6{\cdot}5$. Feel like verifying ? Do that by all means but remember that due to the truncation errors throughtout the computation and rounding off errors at the end, there may not be *absolute* agreement. Instead of getting the right hand sides as 2·5, 25·0 and 47·3, you get respectively the rounded off values 2·5, 25·5 and 47·4.

The above process is known as **Gauss Elimination Method** for obvious reasons. This is a *direct* method of solving a system of linear equations and is quite satisfactory in a large variety of problems. You have already learnt this method in a different guise in Chapter 1. Recall that we could have written the above equations as $\mathbf{AX}=\mathbf{B}$, where

$$\mathbf{A}=\begin{pmatrix} 3 & 1 & -1 \\ 2 & 14 & 3 \\ 5 & 1 & 7 \end{pmatrix},\ \mathbf{X}=\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},\ \mathbf{B}=\begin{pmatrix} 2{\cdot}5 \\ 25{\cdot}0 \\ 47{\cdot}3 \end{pmatrix}.$$

Ultimately we reduce this system to the system of equations (A) which can be written as

$$\begin{pmatrix} 3 & 1 & -1 \\ 0 & 13{\cdot}34 & 3{\cdot}66 \\ 0 & 0 & 6{\cdot}82 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} 2{\cdot}5 \\ 23{\cdot}34 \\ 44{\cdot}28 \end{pmatrix}.$$

Now turn back the pages of the book and have a look at Example 25 in Chapter 1. Compare the solution there and what we have done. Reducing the second and the third elements of the first column to zero amounts to eliminating the first variable from the last two equations and get an equivalent system of equations. The next reduction to zero of the third element in the second column amounts to eliminating the second variable from the last equation of the new system. Thus triangularizing the coefficient matrix is precisely the above process of elimination, only put elegantly. Hence without further ado, we shall now go on to an iterative process for solving systems of linear equations.

15·10·2. Gauss-Seidel Method

Sometimes we have some inkling as to an approximate value of one or more variables in a given system of linear equations. In such cases it may be more economical to use an iterative process where each iteration generates better and better approximations. Indeed one may start with any arbitrary values for the variables, and in case the process is convergent, the solution would eventually be obtained. Let us illustrate one iteration technique which is known as **Gauss-Seidal Method.**

Let us use the same system of equations as we took in case of Gauss elimination method. How we can solve the first (resp. second, third) equation for x_1 (resp. x_2, x_3) as follows :

$$x_1=\frac{2\cdot5-x_2+x_3}{3}=0\cdot33-0\cdot33x_2+0\cdot33x_3, \qquad ...(1)$$

$$x_2=\frac{25-2x_1-3x_3}{14}=1\cdot78-0\cdot14x_1-0\cdot21x_3, \qquad ...(2)$$

$$x_3=\frac{47\cdot3-5x_1-x_2}{7}=9\cdot46--x_1-0\cdot20x_2 \qquad ...(3)$$

The process starts by assuming any values for x_2 and x_3 and computing x_1 from (1) on the basis of these values. We may assume x_1 also to be zero, or any other value, but this would not be used. For example, in the absence of any guidelines, let us take $x_2=x_3=0$. Then an estimate of x_1 from (1) is

$$x_1=0\cdot83.$$

Thus our current values are

$$x_1=0\cdot83,\ x_2=x_3=0.$$

Substituting $x_1=0\cdot83$ and $x_3=0$ in the R.H.S. of (2), we now get an estimate of x_2 as

$$x_2=1\cdot78-0\cdot14\times0\cdot83=1\cdot66.$$

We now replace the old estimate 0 of x_2 by 1·66. Thus we have

$$x_1=0\cdot83,\ x_2=1\cdot66,\ x_3=0.$$

Substituting $x_1=0\cdot83$ and $x_2=1\cdot66$ in the R.H.S. of (3) to estimate x_3, we have

$$x_3=9\cdot46-0\cdot83-0\cdot20\times1\cdot66=8\cdot29.$$

We now replace the old estimate 0 of x_3 by 8·29. Thus after the first iteration, the current solution is

$$x_1=0\cdot83,\ x_2=1\cdot66,\ x_3=8\cdot29.$$

We shall better this solution by using (1), (2) and (3) with the recent-most estimates of the variables being substituting in the right hand sides. Substituting $x_2=1\cdot66$ and $x_3=8\cdot29$ in the R.H.S. of (1),

$$\begin{aligned}x_1&=0\cdot83-0\cdot33\times1\cdot66+0\cdot33\times8\cdot29,\\&=0\cdot83+0\cdot33(8\cdot29-1\cdot66)=3\cdot01.\end{aligned}$$

We now replace the old estimate 0·83 of x_1 by 3·01. Thus

$$x_1=3\cdot01,\ x_2=1\cdot66,\ x_3=8\cdot29.$$

Substituting $x_1=3\cdot01$ and $x_3=8\cdot29$ in the R.H.S. of (2), the new estimate of

$$\begin{aligned}x_2&=1\cdot78-0\cdot14\times3\cdot01-0\cdot21\times8\cdot29,\\&=-0\cdot38.\end{aligned}$$

Replace the old estimate 1·66 of x_2 by $-0·38$. Thus

$$x_1=3·01,\ x_2=-·38,\ x_3=8·29.$$

Substituting $x_1=3·01$ and $x_2=-·38$ in the R.H.S. of (3), the new estimate of

$$x_3=9·46-3·01-0·20\times(-·38),$$
$$=6·38.$$

Hence after the second iteration,

$$x_1=3·01,\ x_2=-·38,\ x_3=6·38.$$

We iterate the process to find a new (and hopefully better !) set of values of x_1, x_2, x_3 using always the recent most values. These are given by

$$x_1=0·83-0·33\times(-·38)+0·33\times6·38=3·06.$$
$$x_2=1·78-0·14\times(3·06)-0·21\times6·38=0·03.$$
$$x_3=9·46-3·06-0·20\times0·03=6·40.$$

Hence the solution after the third iteration is

$$x_1=3·06,\ x_2=0·03,\ x_3=6·40.$$

Iterating again,

$$x_1=0·83-0·33\times0·03+0·33\times6·40=2·93.$$
$$x_2=1·78-0·14\times2·93-0·33\times6·40=0·74.$$
$$x_3=9·46-2·93-0·20\times0·74=6·39.$$

Iterating again,

$$x_1=0·83-0·33\times0·74+0·33\times6·39=2·71.$$
$$x_2=1·78-0·14\times2·71-0·21\times6·39=0·05.$$
$$x_3=9·46-2·71-0·20\times0·05=6·74.$$

Iterating again,

$$x_1=0·83-0·33\times0·05+0·33\times6·74=3·05$$
$$x_2=1·78-0·14\times3·05-0·21\times6·74=0·05$$
$$x_3=9·46-3·05-0·20\times0·05=6·40.$$

Two further iterations produce the following values :

$$x_1=2·92,\ x_2=0·02 \text{ and } x_3=6·54$$
$$x_1=2·98,\ x_2=0·01 \text{ and } x_3=6·48$$

It is clear that the solution has settled down or converged. Rounded off to one decimal place, the solution is given by

$$x_1=3,\ x_2=0, \text{ and } x_3=6·5,$$

which is in perfect agreement with our earlier solution. This need not be the case when greater accuracy is warranted. Compare, *e.g.*, the two solutions before rounding off.

Remark 1. In general, a large number of iterations are required before the solution settles down. The number of iterations,

here as elsewhere, depends upon the data as well as the initial values. The closer the initial values to the solution, the lesser the number of iterations required.

2. It is helpful to go on tabulating the new values in order to see whether the process is converging, when to stop, and which values to use at any stage. We can organize the work of the above example as follows. The 'dashes' in the table indicate that the value above (the dash) is to be used.

I_N	X_1	X_2	X_3
I_0		0	0
I_1	0.83 — —	— 1.66 —	— — 8.29
I_2	3.01 — —	— −0.38 —	— — 6.38
I_3	3.06 — —	— 0.03 —	— — 6.40
I_4	2.93 — —	— 0.74 —	— — 6.39
I_5	2.71 — —	— 0.05 —	— — 6.74
I_6	3.05 — —	— 0.05 —	— — 6.40
I_7	2.92 — —	— 0.02 —	— — 6.54
I_8	2.98 — —	— 0.01 —	— — 6.48

Example 32. *Apply Gauss-Seidel iteration technique to solve the equations*

$$2x+3y=5,$$
$$5x-2y=3.$$

Solution. Perhaps you have solved the equations already orally and found the solution $x=y=1$. Let us choose the initial values $x=0$ and $y=0$ and start the process. We shall always use the most-recent estimates for both x and y. The table below lists the solutions $x=x_n$, $y=y_n$ after the nth iteration. The given equations are written as

$$x=\frac{5-3y}{2}=2{\cdot}5-1{\cdot}5y, \quad \ldots(1)$$

$$y=\frac{5x-3}{2}=2{\cdot}5x-1{\cdot}5 \quad \ldots(2)$$

Substituting $y=0$ in the R.H.S. of (1),

$$x_1=2{\cdot}5.$$

Substituting this estimate of x_1 in the R.H.S. of (2), we get

$$y_1=6{\cdot}25-1{\cdot}5=4{\cdot}75.$$

Hence after the first iteration, the solution is given by

$$x=2{\cdot}5,\ y=4{\cdot}75$$

Iterating again,

$$x_2=2{\cdot}5-1{\cdot}5\times4{\cdot}75=-4{\cdot}625,$$
$$y_2=2{\cdot}5\times(-4{\cdot}625)-1{\cdot}5=10{\cdot}0625.$$

The third iteration now produces,

$$x_3=2{\cdot}5-1{\cdot}5\times10{\cdot}0625=-12{\cdot}5937,$$
$$y_3=2{\cdot}5\times(-12{\cdot}5937)-1{\cdot}5=-32{\cdot}9842.$$

I_n	x_n	y_n
I_0 (initialization)		0
I_1	2.5 —	— 4.75
I_2	−4.625 —	— 10.0625
I_3	−12.5937 —	— −32.9842
⋮	⋮	⋮

What *is* this ? The solution (1, 1) obtained orally is certainly correct. So, have Gauss-Seidel sent us on a wild goose-chase ? Instead of coming any near, we are shooting away from the real solution. *The process does not converge* ; and why should you feel surprised ? What was the guarantee that it would converge at all ?

Here is a greater surprise if you have recovered from the earlier one. Let us use the second equation to estimate x and the first to estimate y. Thus we have

$$x=\frac{2y+3}{5}=0\cdot4y+0\cdot6, \qquad \text{...(3)}$$

$$y=\frac{5-2x}{3}=1\cdot66-0\cdot66x \qquad \text{...(4)}$$

Starting with $x=0=y$ and always using the recent-most values of x, y,

$$x_1=0\cdot6,$$
$$y_1=1\cdot66-0\cdot66\times0\cdot6=1\cdot26.$$
$$x_2=0\cdot4\times1\cdot26+0\cdot6=1\cdot10,$$
$$y_2=1\cdot66-0\cdot66\times1\cdot10=0\cdot93,$$
$$x_3=0\cdot4\times0\cdot93+0\cdot6=0\cdot97,$$
$$y_3=1\cdot66-0\cdot66\times0\cdot97=1\cdot01.$$
$$x_4=0\cdot4\times1\cdot01+0\cdot6=1\cdot00,$$
$$y_4=1\cdot66-0\cdot66\times1\cdot00=1\cdot00.$$

I_n	x_n	y_n
0	0	0
I_1	0.6 —	— 1.26
I_2	1.10 —	— 0.93
I_3	0.97 —	— 1.01
I_4	1.00 —	— 1.00

There ! We have got the solution. Why has this happened ? Let us scrutinize what we did. In the first case, we wrote the first equation as

$$x=2\cdot5-1\cdot5y. \qquad \text{...(1)}$$

When we take an approximate value 0 for the actual value 1 of y, we are in absolute error of 1 unit in y. When we estimate x from (1) with this much error in the value of y, x_1 is in an absolute error of

$1.5 \times 1 = 1.5$. When this value of x with absolute error 1.5 is used to estimate y_1 from the equation

$$y = 2.5x - 1.5, \qquad ...(2)$$

the error in y_1 is 2.5×1.5. Next time, the error in x_2 is $1.5 \times 2.5 \times 1.5$ and that in y_2 is $2.5 \times 1.5 \times 2.5 \times 1.5$. And this mounting of error goes on unchecked. That is why our solution did not converge.

In the second case, the coefficients of x and y (resp. 0.4 and $-.66$) on the right hand sides of the equation (3) and (4) were less than 1 numerically. For this reason, the absolute errors caused by taking the estimates instead of the actual values went of reducing and we ended up with the correct solution.

What is the moral to be drawn from this example ? We should so arrange the equations that the coefficients of the variables appearing on the right hand sides should be less than one numerically. This amounts to dividing by the bigger coefficient. For example. in $2x+3y=5$, y has the bigger co-efficient. So we should use it for getting the successive estimates of y and write it as $y=\cdots\cdots$, *i.e.*, we should *pivot* y in this. In the other equation $5x-2y=3$, x has the bigger coefficient. So we pivot x and write it as $x=\cdots\cdots$ If by chance, this cannot be done, then there is a question-mark on the wisdom of using Gauss-Seidel method. Note that ***bigger*** here means ***bigger in magnitude***. Thus in equation $2x-5y=6$, the coefficient of y is to be treated as bigger and this equation should be used to pivot y.

In case of three variables, we should so organize matters that the magnitude of the coefficient of the pivoted variable is greater than or equal to the sum of the magnitudes of those of the others with at least one inequality being a strict inequality. Analytically, let the given equations be

$$\sum_{i=1}^{3} a_i x_i = d_1, \quad \sum_{i=1}^{3} b_i x_i = d_2, \quad \sum_{i=1}^{3} c_i x_i = d_3.$$

Then a *sufficient* condition for Gauss-Seidel process to converge with x_1, x_2, x_3 being pivoted from the first, the second, and the third equation respectively is that

$$|a_1| \geqslant |a_2| + |a_3|,$$
$$|b_1| \geqslant |b_2| + |b_3|,$$
$$|c_1| \geqslant |c_2| + |c_3|,$$

with at least one of the inequalities being strict. The order of equations is unimportant. For instance we changed the order of the given equations in Example 19 and made the process converge. The condition given above is only *sufficient*. Occasionally, the process may converge even when the condition is not satisfied.

EXERCISE 15 (*m*)

1. For which of the following can we say with confidence that Gauss-Seidel Method will converge ?

(*a*) $5x+2y=5,$
$4x+9y=3.$

(*b*) $12x-5y=-3,$
$5x+11y=4.$

(*c*) $4x-2y+2z=1,$
$2x+9y-6z=2,$
$3x+4y-8z=3.$

(*d*) $5x_1+4x_2+x_3=10,$
$2x_1+5x_2+3x_3=16,$
$7x_1+3x_2+10x_3=22.$

2. Rearrange so that Gauss-Seidel process may converge :

(*a*) $2x_1+3x_2-x_3=4,$
$5x_1+2x_2+2x_3=9,$
$4x_1-2x_2+7x_3=1.$

(*b*) $15x_1+2x_2+20x_3=1,$
$15x_1+16x_2-x_3=1,$
$15x_1+10x_2-2x_3=1.$

3. Solve each of the following systems by Gauss-Seidel Method obtaining the answer correct to two decimal places :

(*a*) $12x_1+5x_2=22,$
$3x_1+19x_2=41.$

(*b*) $5x+3y=6,$
$4x+7y=8.$

(*c*) $4x_1+x_2+2x_3=4,$
$3x_1+8x_2-x_3=20,$
$2x_1-x_2-4x_3=4.$

(*d*) $5x_1-x_2+x_3=10,$
$2x_1+4x_2=0,$
$x_1+x_2+5x_3=-1.$

4. Generalize the sufficient condition given for Gauss-Seidel process to converge in *m* unknowns.

5. Solve the following system of equations by Gauss Seidel algorithm, once taking the initial values as $x_2=x_3=x_4=x_5=100$ and once as $x_2=x_3=x_4=x_5=0$:

$$\begin{aligned}4x_1-x_2&=100\\-x_1+4x_2-x_3&=200\\-x_2+4x_3-x_4&=200\\-x_3+4x_4-x_5&=200\\-x_4+4x_5&=100\end{aligned}$$

6. Re-solve all the problems in Exercise 3 above by taking the following initial values :

(*a*) $x_2=1.$ (*b*) $y=\cdot5.$ (*c*) $x_2=1, x_3=-\cdot5.$
(*d*) $x_2=2=x_3, x_4=-1.$

What difference does it make ?

15·11. NUMERICAL INTEGRATION

The fundamental theorem of integral calculus says that IF there exists a function F such that $F'(x)=f(x)$ for all $x\in[a, b]$, then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

The IF in the above theorem is a strong condition and is not easily met with in real life. Quite often, we come across integrals like $\int_a^b e^{-x^2}\,dx$, $\int_a^b \sin x^2\,dx$ etc. which cannot be expressed in terms of familiar functions. Equally often, integrals of interest may arise from some experimental data. Here, the integrand may not be known as a function f defined on all points of some interval $[a, b]$. Instead, all we may know is the value of the integrand at a finite number of points. Evaluation of such integrals being of practical utility, ways and means have been found to evaluate them unmerically. In this section we shall learn two such techniques. Both are based on the basic fact that $\int_a^b f(x)dx$ represents the area under the curve $y=f(x)$, bounded by the x-axis and the ordinates at $x=a$ and $x=b$.

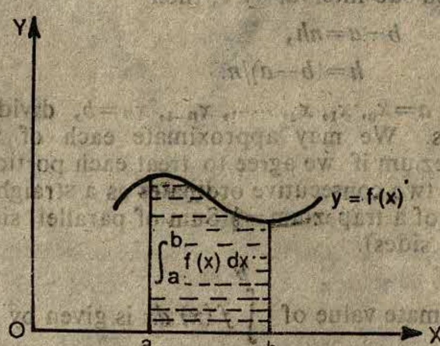


Fig. 15·4.

15·11·1. Trapezium (Trapezoidal) Rule for Numerical Integration

A very simple and elegant method to find an approximate value of $\int_a^b f(x)\,dx$ is the trapezium rule. We divide the interval $[a, b]$ into a finite number, n say, of equal sub-intervals by means of

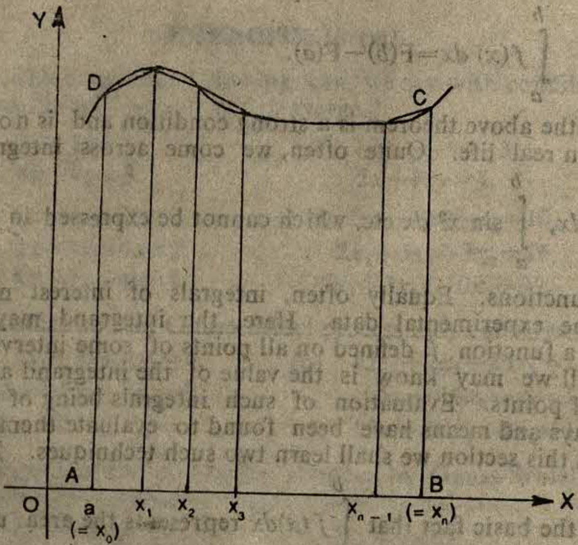


Fig. 15·5.

the points $x_1, x_2, \ldots\ldots, x_{n-1}$ as shown in Fig. 15·5 above. If we denote the length of each sub-interval by e, then

$$b-a=nh,$$

or
$$h=(b-a)/n.$$

The ordinates at $a=x_0, x_1, x_2, \ldots\ldots, x_{n-1}, x_n=b$, divide the required area into n strips. We may approximate each of these strips by means of a trapezium if we agree to treat each portion of the graph trapped between two consecutive ordinates as a straight line segment. Recall that area of a trapezium$=\frac{1}{2}$(Sum of parallel sides)$\times$(distance between parallel sides).

Thus an approximate value of $\int_a^b f(x)\,dx$ is given by

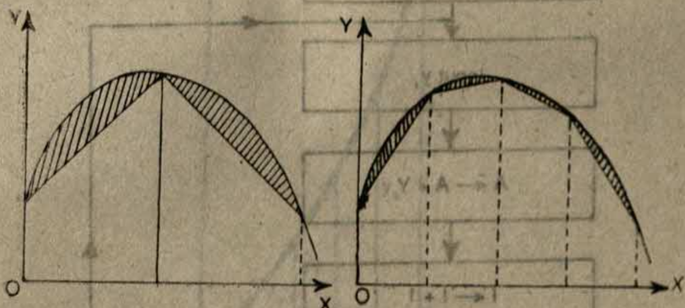
$$\int_a^b f(x)\,dx = \text{area ABCDA} = \tfrac{1}{2}(f(a)+f(x_1))h+\tfrac{1}{2}(f(x_1)+f(x_2))h+\ldots\ldots+\tfrac{1}{2}(f(x_{n-1})+f(b))h,$$

$$=\frac{h}{2}\Big[f(a)+f(b)+2\Big\{f(x_1)+f(x_2)+\ldots\ldots+f(x_{n-1})\Big\}\Big].$$

$$=h[\tfrac{1}{2}(f(a)+f(b))+(f(x_1)+f(x_2)+ \cdots\cdots +f(x_{n-1}))] \quad \cdots(A)$$

The above formula for an approximate value of $\int_a^b f(x)\,dx$ is known as the **trapezium** or **trapezoidal rule** for evaluating definite integrals approximately.

Remarks 1. Are you wondering how many points $x_1, x_2, \cdots, x_n$ should one take ? Well ! That depends on how much accuracy you want. Let us observe that in general, the more the number of strips, the lesser is the error (shown dotted) in treating the sum of the trapezia as the area under the graph. Hence the more the number of points, the greater is the accuracy. Generally we go on doubling the number of sub-intervals and compute the area. When two successive approximations to the area agree upto the required accuracy, we stop.



2. It is not necessary to take the lengths of the sub-intervals equal. Equal sub-intervals are taken for the sake of simplicity alone.

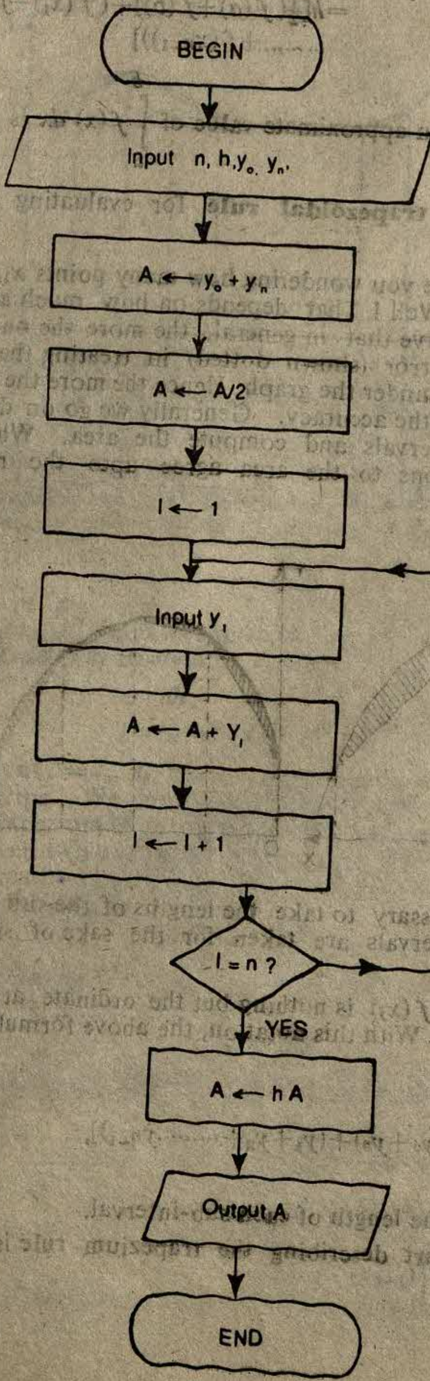
3. Recall that $f(x_j)$ is nothing but the ordinate at x_j. We may as well call it y_j. With this notation, the above formula can be written as

$$\int_a^b f(x)\,dx=h[\tfrac{1}{2}(y_0+y_n)+(y_1+y_2+\cdots\cdots y_{n-1})],$$

where $h=(b-a)/n$ is the length of each sub-interval. $\quad\cdots(5)$

4. The flow-chart describing the trapezium rule is given on page 880.

Flow-chart Describing Trapezium-rule

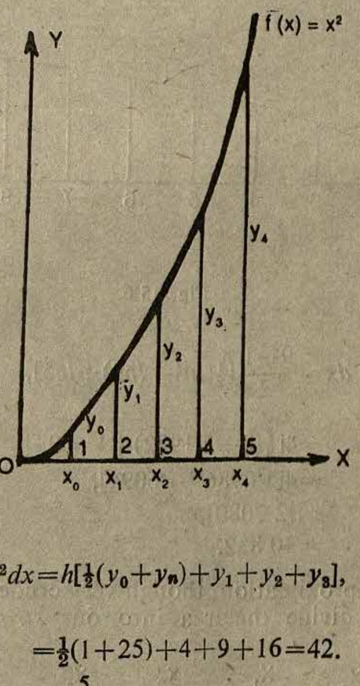


Example 33. *Evaluate* $\int_1^5 x^2\,dx$ *using the trapezium rule taking* $n=4$.

Solution. The range of integration [1, 5] has length 4.

Taking n (number of strips)$=4$, $h=\frac{(b-a)}{n}=\frac{(5-1)}{4}=1.$

Hence $x_0=1$, $x_1=2$, $x_2=3$, $x_3=4$, $x_4=5$. Also $y_0=f(x_0)=f(1)=1$, $y_1=f(2)=4$, $y_2=f(3)=9$, $y_3=f(4)=16$, $y_4=f(5)=25$.



$$\therefore \quad \int_1^5 x^2 dx = h[\tfrac{1}{2}(y_0+y_n)+y_1+y_2+y_3],$$

$$=\tfrac{1}{2}(1+25)+4+9+16=42.$$

Directly evaluating $\int_1^5 x^2 dx$, we get

$$\int_1^5 x^2\,dx = 0{\cdot}33\, x^3 \Big|_1^5 = 0{\cdot}33(125-1)$$

$$=0{\cdot}33\times 124$$
$$=40{\cdot}92.$$

Remark. The above example was chosen for ease of demonstration. The real worth of the rule lies where we are forced to integrate numerically.

Example 34. *Evaluate* $\int_1^9 \ln x\, dx$ *by the trapezium rule, taking* $n=2$, *4 and 8, one by one.*

Solution. The graph of the function is shown in Fig. 15·6. Let us first take $n=2$, *i.e.*, let us divide the area into two strips by means of a point x_1 mid-way between 1 and 9. This point is 5. Now using the trapezium rule.

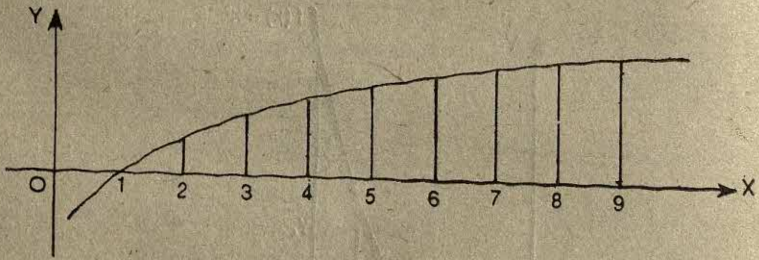


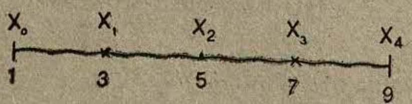
Fig. 15·6.

$$\int_1^9 \ln x\, dx = \frac{9-1}{2}[\tfrac{1}{2}(\ln 1+\ln 9)+\ln 5],$$

$$=4[\tfrac{1}{2}(0+2{\cdot}1972)+1{\cdot}6094], \qquad \ldots(1)$$
$$=4[1{\cdot}0986+1{\cdot}6094],$$
$$=[2{\cdot}7080],$$
$$=10{\cdot}832.$$

This is the first approximation, though very crude.

Let us now divide the area into four *strips* by means of the



three evenly-spaced points $x_1=3$, $x_2=5$ and $x_3=7$ between 1 and 9. Note that the additional points are the middle points of the earlier sub-intervals [1, 5] and [5, 9]. Again using the trapezium rule,

$$\int_1^9 \ln x\, dx = \frac{9-1}{4}[\tfrac{1}{2}(\ln 1+\ln 9)+(\ln 3+\ln 5+\ln 7)],$$

$=2[2{\cdot}7080+(ln3+ln7)],$ (using the value 2·7080 of

$=2[2{\cdot}7080+(1{\cdot}0986+1{\cdot}9459)],$ $\frac{1}{2}(ln1+ln9)+ln5$ computed earlier.)

$=2[5{\cdot}7525],$

$=11{\cdot}5050.$

This is the second approximation.

Let us now divide the area into *eight* strips. This calls for considering the middle points 2, 4, 6 and 8 of the earlier sub-intervals. The additional term inside the brackets is $(ln2+ln4+ln6+ln8)$. Using the earlier computations,

$$\int_1^9 lnx\,dx=\frac{9-1}{8}[5{\cdot}7525+ln2+ln4+ln6+ln8],$$

$$=5{\cdot}7525+5{\cdot}9506$$

$$=11{\cdot}7031.$$

Remarks. 1. For better values, we may increase the number of strips further.

2. By direct evaluation (intrgrating by parts with $ln\ x$ as the first function and 1 as the second),

$$\int_1^9 lnx\,dx=9ln9-8$$

$$=19{\cdot}7750-8$$

$$=11{\cdot}775.$$

Hence the last approximate value is in absolute error

$$=11{\cdot}775-11{\cdot}7031,$$

$$=0{\cdot}0719.$$

The per cent relative error is thus $07{\cdot}19/11{\cdot}775=0{\cdot}61$ nearly.

The trapezium rule generally produces good results with a moderate number of points of subdivision.

3. If you do not have a pocket calculator, make use of the (common) log tables, using the suitable multiplication factor. Recall that $ln\ x=ln\ 10\ \log x=2{\cdot}3025\log x$.

4. We can organize the work above, for $n=4$ say, in a compact form as shown below. The values of $f(a)\ (=f(x_0))$ and $f(b)\ (=f(x_n))$ are kept in a separate column. All other $f(x_j)$'s are kept in a different column.

Evaluation of $\int_1^9 \ln x \, dx$ *by the Trapezium Rule*

n	x_n	$\ln x_n$	
0	1 $(=x_0=a)$	0	
1	3		1.0986
2	5		1.6094
3	7		1.9459
4	9 $(=x_4=b)$	2.1972	
Totals		2.1972	4.6539

$$\text{Hence} \int_1^9 \ln x \, dx = \frac{b-a}{4}[\tfrac{1}{2}\{(f(a)+f(b)\}+\{f(x_1)+f(x_2)+f(x_3)\}],$$

$$=2[\tfrac{1}{2}\times 2{\cdot}1972+4{\cdot}6539]$$

$$=11{\cdot}5050.$$

Example 35. *Some values of a certain function f are tabulated below* :

X	1 0	1.1	1 2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
f(x)	1 543	1 668	1.811	1.971	2.151	2.352	2.577	2.828	3.107	3.417	3.762

Find $\int_1^2 f(x)\, dx$ *using trapezoidal rule with* $h=\cdot 2$.

Solution. Length of the interval [1, 2] is 1. If each sub-interval is to be of length 0·2, then there are to be five sub-intervals in all. Thus, $n=5$. The points of division are 1·2, 1·4, 1·6 and 1·8 in addition to the end-points 1 and 2. The relevant data are listed in the following table :

Evaluation of $\int_1^2 f(x)\,dx$ *by the Trapezium Rule*

n	x_n	$f(x_n)$	
		$f(x_0)$, $f(x_n)$	$f(x_1),\ldots,f(y_{n-1})$
0	1.0	1.543	
1	1.2		1.811
2	1.4		2.151
3	1.6		2.577
4	1.8		3.107
5	2.0	3.762	
Totals		5.305	9.646

$$\int_1^2 f(x)\,dx=\frac{2-1}{5}[\tfrac{1}{2}(5{\cdot}305)+9{\cdot}646],$$

$$={\cdot}2(2{\cdot}6525+9{\cdot}646)$$
$$=2{\cdot}4597$$

Remark. f above is defined by $f(x)=(e^x+e^{-x})/2$. The actual value of the integral truncated to two places of decimal is 2·45.

15·11·2. Simpson's Rule for Numerical Integration

The trapezoidal rule approximates the function f by a linear function each over the various sub-intervals. However, most functions that we come across are curvilinear rather than linear. Hence there is every chance of reducing error if we approximate f by means of a second degree curve—a suitable parabola in particular, over small portions of the graph. Since a parabola can be uniquely determined by three points, we shall combine our sub-intervals in pairs, every pair of adjacent sub-intervals giving us three points on the graph This requires that we *sub-divide the interval* $[a, b]$ *into an even number of sub-intervals* by means of the points

$$a=x_0,\ x_1,\ x_2,\ \cdots\cdots,\ x_{2n-1},\ x_{2n}=b \text{ say.}$$

Let the ordinates $y_0, y_1, y_2, \cdots\cdots, y_{2n}$ at these points meet the graph at the points $P_0, P_1, P_2, \cdots\cdots, P_{2n}$.

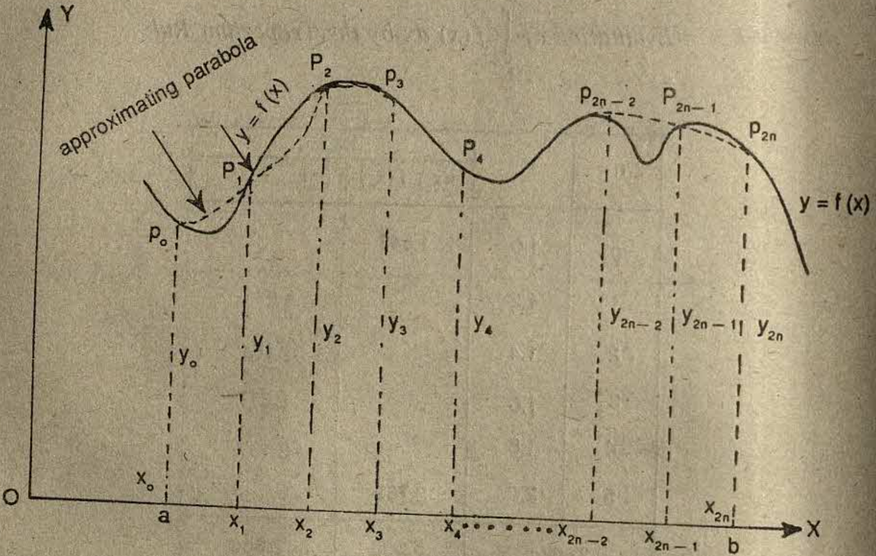


Fig. 15·7.

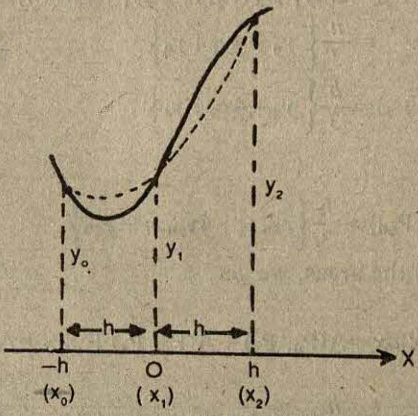
We shall first take P_0, P_1 and P_2, and fit a parabola passing through the same, the axis of the parabola being taken vertical. We shall then approximate the graph of f between P_0 and P_2 by means of this parabola. In effect, the area enclosed by the fitted parabola, the x-axis and the ordinates y_0 and y_2, would be taken as equal to the area enclosed by the graph of f, the x-axis and the ordinates y_0, y_2. Next, we shall take P_2, P_3 and P_4; approximate the graph of f between P_2 and P_4 by a new parabola, and compute the area under this parabola between P_2 and P_4; and so on. Ultimately, the areas under the various parabolas would be summed up and the sum taken as an approximate value of the required area.

Let us now see how to fit a parabola passing through three given points P_0, P_1, P_2 and what is the area, A say, under this parabola between P_0 and P_2. Notice first of all that we are interested in the value of A. If we were to slide this area A (to the left or right) till the ordinate y_1 coincides with the y-axis, no change would occur in the value of A. So let us assume for the moment that the origin is at x_1 and y_1 lies along the y-axis.

Since P_0, P_1, P_2 are given, the ordinates y_0, y_1, y_2 at them are, therefore, known quantities. Suppose that the equation of the required parabola is

$$y=p+qx+rx^2. \quad \text{...(1)}$$

We have to determine p, q and so that the parabola (1), shown dotted in Fig. 15·7, may pass through P_0, P_1 and P_2. If the length of



each sub-interval is h, then $x_0=-h$ and $x_2=h$. P_0, P_1 and P_2 thus become the points $(-h, y_0)$, $(0, y_1)$ and (h, y_2) respectively. Since the parabola passes through these points, we must have

$$y_0=p-qh+rh^2, \qquad \text{...(2)}$$

$$y_1=p, \quad \text{and} \qquad \text{...(3)}$$

$$y_2=p+qh+rh^2. \qquad \text{...(4)}$$

Also, denoting the area under the parabola between P_0 and P_2 by $A(P_0, P_2)$,

$$A(P_0, P_2)=\int_{-h}^{h} y\,dx=\int_{-h}^{h} (p+qx+rx^2)\,dx,$$

$$=\Big| px+\tfrac{1}{2}qx^2+\tfrac{1}{3}rx^3 \Big|_{-h}^{h},$$

$$=2(ph+\tfrac{1}{3}rh^3). \qquad \text{...(5)}$$

In the above expression (5) for the area, p and r are the unknown quantities, whereas since P_0, P_1, P_2 are given points, the ordinates y_0, y_1, y_2 are known. (As a matter of fact, $y_0=f(x_0)$, $y_1=f(x_1)$ and $y_2=f(x_2)$.) We shall, therefore, use equations (2), (3) and (4) to determine p and r. From (3),

$$p=y_1. \qquad \text{...(6)}$$

Adding (2) and (4),

$$y_0+y_2=2p+2rh^2,$$

or $$y_0+y_2=2y_1+2rh^2, \qquad \text{[From (6)]}$$

or $$rh^2=\tfrac{1}{2}(y_0-2y_1+y_2). \qquad \text{...(7)}$$

Substituting the values of p and rh^2 from (6) and (7) in (5), we get

$$A(P_0, P_2)=2\{hy_1+\tfrac{1}{3}\times\tfrac{1}{2}(y_0-2y_1+y_2)\times h\},$$
$$=\frac{h}{3}\left\{y_0+4y_1+y_2\right\}.$$

Similarly, $A(P_2, P_4)=\dfrac{h}{3}\left\{y_2+4y_3+y_4\right\}$,

$$\vdots \qquad\qquad \vdots$$

$$A(P_{2n-2}, P_{2n})=\frac{h}{3}\left\{y_{2n-2}+4y_{2n-1}+y_{2n}\right\}.$$

Summing up all the areas, we get,

$$\int_a^b f(x)dx=A(P_0, P_2)+A(P_2, P_4)+\ldots\ldots+A(P_{2n-2}, P_{2n}).$$

$$=\frac{h}{3}(y_0+4y_1+y_2)+(y_2+4y_3+y_4)+\ldots\ldots$$
$$\ldots\ldots+(y_{2n-2}+4y_{2n-1}+y_{2n}),$$
$$=\frac{h}{3}\Big[(y_0+y_{2n})+2(y_2+y_4+\ldots\ldots+y_{2n-2})$$
$$+4(y_1+y_3+\ldots\ldots+y_{2n-1})\Big],$$

as an approximate value of the area under $y=f(x)$ between the ordinates at a and b. This formula is known as the *Parabolic Rule* or *Simpson's One-third Rule* or simply *Simpson's Rule* for numerical integration. The formal description of the rule is given below :

Step 1. Divide $[a, b]$ into $2n$ sub-intervals by $(a=)\ x_0, x_1, x_2, \ldots\ldots, x_{2n}\ (=b)$, for a suitable n.

Step 2. Compute $y_0, y_1, \ldots\ldots, y_{2n}$.

Step 3. Form the sum $S_1=y_0+y_{2n}$ of the first and the last ordinates (*i.e.*, the ordinates at a and b).

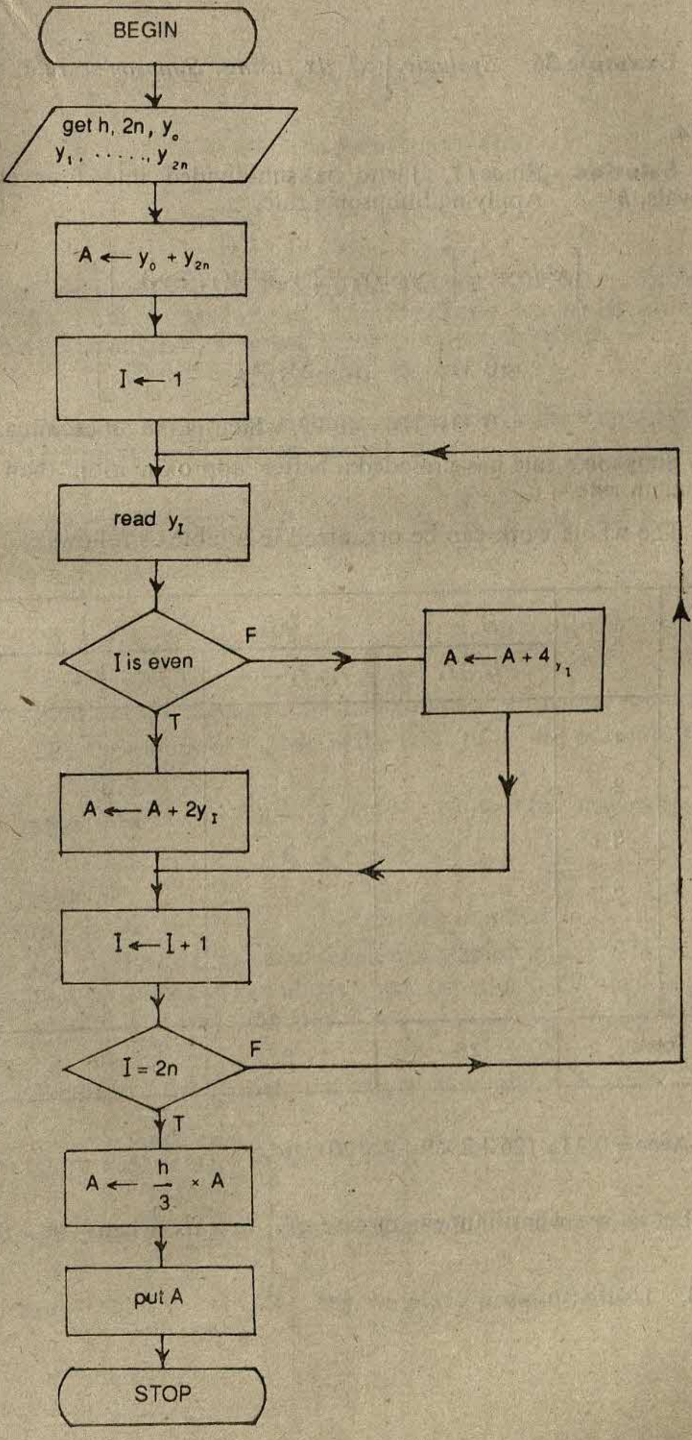
Step 4, Form the sum S_2 of the remaining ordinates with an even suffix (*i.e.*, let $S_2=y_2+y_4+\ldots\ldots+y_{2n-2}$).

Step 5. Form the sum S_3 of all the ordinates with an odd suffix (*i.e.*, let $S_3=y_1+y_3+\ldots\ldots+y_{2n-1}$).

Step 6. Form the sum $S=S_1+2S_2+4S_3$.

Step 7. Multiply S by $\dfrac{h}{3}$. The product $\dfrac{h}{3}\times S$ is an approximate value of $\displaystyle\int_a^b f(x)\,dx$.

The corresponding flow chart is given on page 889.


BEGIN
get h, 2n, y_0
$y_1, \ldots, y_{2n}$
$A \leftarrow y_0 + y_{2n}$
$I \leftarrow 1$
read y_I
I is even
F
T
$A \leftarrow A + 4y_I$
$A \leftarrow A + 2y_I$
$I \leftarrow I + 1$
$I = 2n$
F
T
$A \leftarrow \frac{h}{3} \times A$
put A
STOP

Example 36. *Evaluate* $\int_1^5 x^2\,dx$ *using Simpson's rule with* $2n=4$.

Solution. Since [1, 5] is to be sub-divided into four equal intervals, $h=1$. Applying Simpson's rule,

$$\int_1^5 x^2 dx = \frac{1}{3}\Big[\ y_0+y_4+2(y_2)+4(y_1+y_3)\ \Big],$$

$$=0\cdot 33\Big[\ 1^2+5^2+2\times 3^2+4(2^2+4^2)\ \Big],$$

$$=0\cdot 33\times 124=40\cdot 92, \text{ which is the exact value.}$$

Thus Simpson's rule has provided a better approximation than the trapezium rule.

The whole work can be organized in a table as follows :

n	x_n	y_n		
		y_0 , y_n	y_2 , y_4	y_1 , y_3 ....
0	1	1		
1	2			4
2	3		9	
3	4			16
4	5	25		
Totals		26	9	20

Area$=0\cdot 33\times(26+2\times 9+4\times 20)$ etc.

Let us see what happens in case of $\int_1^9 ln\ x\,dx$ when we take $2n=4$. Using Simpson's rule, we get

$$\int_1^9 ln\, x\, dx=\frac{2}{3}\Big[\,(y_0+y_4)+2(y_2)+4(y_1+y_3)\,\Big],$$

$$=0{\cdot}66[(ln1+ln9)+2ln5+4(ln3+ln7)],$$
$$=0{\cdot}66[2{\cdot}1972+3{\cdot}2188+4\times 3{\cdot}0445)],$$
$$=0{\cdot}66[5{\cdot}4160+12{\cdot}1780],$$
$$=0{\cdot}66\times 17{\cdot}594=11{\cdot}612.$$

Trapezium rule provided 11·5050 for 4 strips. The actual value was 11·775. Thus the area obtained by Simpson's rule is less in error. In fact, by and large, Simpson's rule gives better results.

EXERCISE 15 (*n*)

1. Evaluate each of the following integrals by the trapezium rule for the value of n given against each :

(*a*) $\displaystyle\int_1^2 \frac{1}{x}\,dx,\ n=5.$ (*b*) $\displaystyle\int_0^1 \frac{2}{1+x^2}\,dx,\ n=5.$

(*c*) $\displaystyle\int_0^1 \sqrt{(1-x^2)}\,dx,\ n=4.$ (*d*) $\displaystyle\int_0^{x/2} \sin x\,dx,\ n=5.$

Find out the actual values and compare with your results.

2. Use the values of f given in the table of solved example 35 on page 884 to evaluate $\displaystyle\int_{1\cdot 0}^{1\cdot 8} f(x)dx,$ using the trapezium rule with

(*a*) $h=0{\cdot}1,$ (*b*) $h=0{\cdot}2,$ and

(*c*) $h=0{\cdot}4.$ If the actual value of the integral is 1·7669, then find the errors in the above approximations. Which approximation is in error the least ?

3. Evaluate $\displaystyle\int_{1\cdot 8}^{3\cdot 4} e^x\,dx$ with trapezium rule, taking 8 strips. Find the actual value. What is the error ?

4. Evaluate $\displaystyle\int_0^1 \frac{dx}{1+x}$ using the trapezium rule with h equal to

(*a*) 0·5, (*b*) 0·25, (*c*) 0·125.

Also find the actual value. What are the absolute errors ?

5. Do problem (4) above but use Simpson's rule. Which rule gives better approximations ?

6. Verify that Simpson's rule gives correct values for each of the following with, say four, strips :

(a) $\int_0^1 x^2\,dx.$ (b) $\int_0^1 x^3\,dx.$

7. Using the following table, show that $\int_0^1 e^{-x^2}\,dx = 0{\cdot}7462$ by an application of Simpson's rule :

X	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
e^{-x^2}	1.0	0.99	0.96	0.91	0.85	0.78	0.70	0.61	0.53	0.44	0.37

SUMMARY

Computers operate on data according to their types. Data can be classified as alphabetic, numeric and alphanumeric. Numeric data are of two types—integer and real. Any operations on integer data always produce integer output. Integer operations are simple and faster than the corresponding operations on real data. To enhance the range of numbers which can be represented on a given computer system, we use floating point representations of real numbers. Data cannot always be stored exactly in a computer. For this reason, the input may contain some representational errors also in addition to the inexact measurement errors the input might already involve. These are known as input errors. Some errors are incurred while the operations are being carried out on the data. Thus errors may be caused because of inexact data, inexact operations, or both. Data are said to be numerically instable when a slight change in input causes a drastic change in output. When such a phenomenon occurs because of the form of the input, the instability is called inherent. When this phenomenon occurs because of the operational algorithms, the instability is called induced. Induced instability can be checked but not the inherent one. Numerical methods are concerned with finding approximate solutions to problems. Computers often obtain numerical solutions by the repetition of a certain process (depending upon the problem at hand) a large number of times. Such techniques of solving problems are known as iterative techniques. Where solutions depend upon summing an infinite series to a finite number of terms, it is useful to express the nth term T_n in terms of the previous term T_{n-1}. This has been demonstrated for e^x, sin x and cos x.

Floating point representation :

Manner of writing real numbers with decimal/binary point being placed at the desired place, the adjustment in the value being made by multiplication with a suitable power of ten/two.

Normalized floating point number

A number written as $\cdot x_1x_2\ldots\ldots \times 10^n$ (or $\cdot x_1x_2\ldots\ldots En$) where $x_1, x_2, \ldots\ldots$ are digits, $x_1 \neq 0$, n is any integer. (n is called the **exponent** and $\cdot x_1x_2\ldots\ldots$ is known as the **mantissa**).

If x denotes an approximate value and X the actual value, then

Error$=x-X$

Absolute error$=|x-X|$

Relative error$=|x-X|/X$

Method of bisection

To find a root of $f(x)=0$, find values a and b such that $f(a)\,f(b)<0$. $c=0{\cdot}5(a+b)$ is an approximate root. Replace one of a and b by d so that $f(c)\,f(d)<0$. A root lies between c and d. $0{\cdot}5(c+d)$ is a better approximation. Repeat the process till the root is determined to the desired accuracy.

Method of false position

If a root of $f(x)=0$ lies between a and b, the first approximation c to the root is taken as $a-\dfrac{b-a}{f(b)-f(a)}\,f(a)$. Successive approximations are obtained by modifying the values of a and b as in the method of bisection.

Newton-Raphson method

The initial approximation x_0 must be taken close to the root. The sought out root should be away from the other roots and extreme values. Successive approximations are computed by

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}.$$

Gauss elimination method

It is used to solve a system of m linear equations in m unknowns, First equation is used to eliminate the first variable from the rest of the equations. The second equation in its modified form is used to eliminate the second variable from the remaining equations and so on till the last equation contains just one variable. The values of the various unknowns are now computed by back substitutions.

Gauss-Seidel method

It is an iterative method to solve a system of m equations in m unknowns. It is not always applicable. It begins with writing the given equations in the form

$$x_1=\ldots\ldots,\ x_2=\ldots\ldots,\ \ldots\ldots,\ x_m=\ldots\ldots$$

Using these relations, first x_1 is computed by choosing arbitrary values for $x_2, x_3, \ldots\ldots, x_m$. Then x_2 is computed for the choosen values of $x_3, x_4, \ldots\ldots, x_m$ and the computed value of x_1. Now x_3 is computed from $x_4, \ldots\ldots, x_m$ and new x_1, x_2 and so on. Once the cycle is complete with x_m, we begin again with computing a new value of x_1. All the time, the most recent values are used. Process is terminated when a solution with desired accuracy is obtained.

Trapezium (trapezoidal) rule

This is used to obtain an approximate value of a definite integral $\int_a^b f(x)dx$. The interval $[a, b]$ is divided into n sub-intervals of length

$$h=\left(\frac{b-a}{n}\right)$$

each by the points $a=x_0, x_1, x_2, \ldots\ldots, x_n=b$. With $y_j=f(x_j)$,

$$\int_a^b f(x)\,dx=h[\tfrac{1}{2}(y_0+y_n)+(y_1+y_2+\ldots\ldots+y_{n-1})]$$

approximately.

Simpson's $\left(\frac{1}{3}\text{rd}\right)$ Rule

This time $[a, b]$ is divided into an even number $2n$ of sub-intervals. With the above notation, an approximate value of $\int_a^b f(x)\,dx$ is given by the relation

$$\int_a^b f(x)dx=\frac{h}{3}\Big[(y_0+y_n)+2(y_2+y_4+\ldots\ldots+y_{2n-2})$$

$$+4(y_1+y_3+\ldots\ldots+y_{2n-1})\Big].$$

TEST YOUR UNDERSTANDING XV

1. Which one is not in the normalized floating point form ?
   (*a*) 0·1234E5, (*b*) ·0123E4, (*c*) ·1234E0.
2. $A=2^{10}$, $B=1{\cdot}2345$ and $C=\sqrt{2}$. If in a machine there is provision for a four-digit mantissa and a two-digit exponent (sign apart), then the ones which CANNOT be stored *exactly*, are
   (*a*) A and B, (*b*) B and C, (*c*) A and C.
3. In the normalized floating point representation of ·056789, the mantissa would be
   (*a*) ·056789, (*b*) ·5678, (*c*) ·56789.
4. Using four-digit normalized mantissas, the sum of first 10,000 terms of the series $1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\ldots\ldots$ is found to be the number m. The sum of the first 10,002 terms of this series (in the same mode) would be
   (*a*) greater than m, (*b*) equal to m, (*c*) less than m.
5. If the expressions

$$A=\frac{1}{20^2}+\ldots\ldots+\frac{1}{19^2}+\ldots\ldots\frac{1}{2^2}+1$$

and $$B=1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\ldots\ldots+\frac{1}{400}$$

are each evaluated from left to right using four-digit floating point arithmetic, then
   (*a*) A>B, (*b*) A=B, (*c*) A<B.

6. The unit's digit in the 9's complement of the number 376×245 is
(a) 3, (b) 4,
(c) 0, (d) 9.

7. For any positive integer p, the unit's digit in the 9's complement of $p \times 11$ is
(a) 8,
(b) unit's digit of p,
(c) 9's complement of the unit's digit in p,
(d) not necessarily any of these.

8. If the unit's digit in the 9's complement of p is j, and the unit's digit in the 10's complement of p is k, then
(a) k is always greater than j,
(b) $k=j+1$ in fact,
(c) k is less than j for exactly one value of p,
(d) k is less than j for exactly one value of p.

9. m and n respectively are the number of digits in the 9's and the 10's complement of a positive integer.
(a) $m=n$ *always*,
(b) $n=m+1$ sometimes,
(c) nothing can be said about the relation between m and n,
(d) none of these.

10. 98765·4 truncated to four digits is
(a) 9877, (b) 9876,
(c) 98765, (d) 765·4.

11. 37·5945 round off to 2 decimal places is
(a) 38, (b) 37,
(c) 37·59, (d) 0·59.

12. 37500 rounded off to 2 significant digits is
(a) 37, (b) 38,
(c) 37000, (d) 38000.

13. 37·650 rounded off to one decimal place is
(a) 37·6, (b) 37·7,
(c) 0·6, (d) 0·7.

14. Truncating a seven-digit number to four digits amounts to
(a) removing the four least significant digits,
(b) removing the three least significant digits,
(c) replacing the three least significant digits by zeros,
(d) none of these.

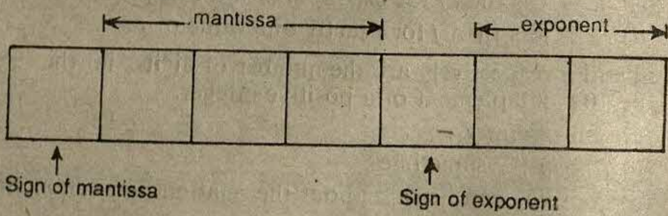
15. If f is a function continuous on $[a, b]$ and $f(a)\ f(b)<0$, then $f(x)=0$ must have a root
   (*a*) between a and b,
   (*b*) equal to one of a and b,
   (*c*) less than a or greater than b.

16. If f is continuous on $[a, b]$, and $f(a)\,f(b)<0$, then the number of roots of $f(x)=0$ which lie between a and b CANNOT be
   (*a*) 1, (*b*) 2, (*c*) 3.

REVIEW EXERCISE XV

1. What would be the exponent when 376·376E−5 is written with 0·376376 as the mantissa ?

2. What answers would be obtained on a machine which stores numbers as follows :



   (*a*) $0·34789+25·65210$. (*b*) $0·34789-25·65210$.
   (*c*) $123\times 0·123$. (*d*) $0·256\times 256$.

3. Prove that an error bound for a floating point addition operation is 5×10^{-4} times the result, when rounding is done to get a four-digit mantissa. Deduce that addition involving more than two summands should be carried out in ascending order of summands.

4. The equation whose roots are $-1, -2, -3, \ldots\ldots, -20$, is $x^{20}+210x^{19}+\ldots\ldots+20\,!=0$. Now a very small factor 2^{-23} (it has six consecutive zeros after the decimal point) is added to the co-efficient of x^{19}. The new equation is found to have ten complex roots. What kind of instability is that—inherent or induced ?

5. Starting with a suitable initial approximation x_0, find $\sqrt{180}$ correct to decimal places by the iterative process

$$x_{n+1}=0·5\left(\frac{180}{x_n}+x_n\right).$$

6. Evaluate e rounded off to two significant digits by the use of some iterative process.

7. The equation $x^5-5x+1=0$ has one negative and two positive real roots. Determine intervals of length one each which contain these roots.

8. Use the method of bisection to find the positive root of $x^2-2=0$ to four decimal places. Use Newton-Raphson's method to do the same.

9. Use Newton-Raphson's algorithm to find a root of $x^5+5x+1=0$ truncated at four decimal places.

10. A root of the equation $\cos x-3x+1=0$ lies near 0·5. Use three different methods to evaluate this root to four decimal places.

11. Make use of the following table and show by the trapezium rule that $\int_0^{0 \cdot 8} e^{x^2}\,dx=1{\cdot}0192$.

X	0.0	0.2	0.4	0.6	0.8
e^{x^2}	1.0	1.0408	1.1735	1.4333	1.8965

12. Evaluate $\int_3^7 x^2 \log x\,dx$ using Simpson's rule with four strips.

13. Solve using Gauss elimination method the following equations :

$$5x+2y+10z=44,$$
$$x+5y-2z=3,$$
$$10x-y+3z=29.$$

14. Re-organize and solve the system of equations given in the previous problem by Gauss-Seidel Method.

15. Which of the following statements are true ?

(*a*) The number of significant digits in the 9's complement of m is the same as that in m.

(*b*) The number of significant digits in the 10's complement of m is the same as that in m.

(*c*) The mantissas of $\sqrt{0{\cdot}3}$, $\sqrt{3}$ and $\sqrt{30}$ are the same in the normalized floating point representation.

(*d*) Gauss-Seidel method can be used to solve any three linear equations in three unknowns.

HISTORICAL NOTE

In some sense, the history of numerical methods begins with the development of logarithms in the 16th century. The age of Newton saw the infant stages of iterative methods. What is now known as Newton-Raphson method for finding a root of $f(x)=0$ was used in its most primitive form by Newton to solve a cubic equation. With the passage of time, it has established itself as a popular technique to solve the most general of the equations of the type $f(x)=0$. Euler, Lagrange and Laplace are other names which must be remembered, though in this introductory discourse on numerical methods we have not been able to touch upon what they did. The next name is that of Gauss. He was the first to treat the subject of rounding errors in a systematic way in his work **Theoria Motus**. In this book itself, Gauss treats the subject of numerical instability also. Gauss did a lot of work on numerical integration as well. Gauss gave several techniques of solving systems of equations. What is now known as Gauss-Seidel method was described in a crude form to Gerling by Gauss. Gerling published it in 1845 in a work on practical geometry. Jacobi also published a similar method at about the same time. Ludwig Seidel was a student of Jacobi. He improved upon the technique; hence the name. Do you know that Thomas Simpson (1710—1761) was a silk-weaver by trade? Mathematics was kind of a hobby with him. He wrote several books on mathematics which ultimately earned him a Professorship at the Military College in Woolwich, U.K. Floating point representations and operations are of course, a sequel to the modern day computers.

□□

ANSWERS

उपदेशलवं शास्त्रं कुरुते धीमतो यतः ।
तत्तु प्राप्यैव विस्तार स्वयमेवोपगच्छति ॥
जले तैलं खले गुह्यं पात्रे दानं मनागपि ।
प्राज्ञे शास्त्रं स्वयं याति विस्तारं वस्तुशक्तितः ॥

★ ★ ★

A little instruction and guidance in science is sufficient for the intelligent student, for this alone will help him to develop his knowledge of his own accord. Science instilled into the intelligent mind has sufficient vitality in it to grow and expand by its own force like a drop of oil on a sheet of water, a piece of secret confined to a villain, or a little act of charity to the deserving person.

—Bhaskara

ANSWERS

[*For each of the answers marked with an asterisk* (*), *an example different from the one given below could also have served the purpose*].

Exercise 1 (*a*)

1. (*a*) $\begin{pmatrix} 3 & 1 \\ 3 & 3 \end{pmatrix}$. (*b*) $\begin{pmatrix} 10 & 9 & 9 \\ 8 & 2 & 13 \end{pmatrix}$.

2. $\begin{pmatrix} 3 & 6 & 18 \\ 9 & -3 & 12 \end{pmatrix}$, $\begin{pmatrix} -4 & -8 & -24 \\ -12 & 4 & -16 \end{pmatrix}$,

$\begin{pmatrix} -1 & -2 & -6 \\ -3 & 1 & -4 \end{pmatrix}$.

3. $\begin{pmatrix} -1 & 1 & -4 \\ -2 & -8 & -9 \end{pmatrix}$.

4. $\begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1 & 1/2 & 3/4 \\ -1/4 & 9/4 & 7/4 \end{pmatrix}$. **6.** $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Exercise 1 (*b*)

1. (*a*) $\begin{pmatrix} 0 & 2 & -1 \\ -6 & -5 & 6 \\ 3 & 6 & -4 \end{pmatrix}$ (*b*) $\begin{pmatrix} -8 & -5 & 1 \\ 8 & 4 & 4 \\ 4 & 4 & 1 \end{pmatrix}$

(*c*) $\begin{pmatrix} 6 & -3 & 5 \\ -2 & 5 & -7 \\ 2 & -1 & 2 \end{pmatrix}$ (*d*) $\begin{pmatrix} -2 & -5 & 7 \\ 3 & -2 & 3 \\ 4 & 4 & -5 \end{pmatrix}$

(*e*) $\begin{pmatrix} 6 & 9 & -2 \\ -6 & -11 & 6 \\ 3 & 0 & 2 \end{pmatrix}$ (*f*) $\begin{pmatrix} 4 & 2 & 1 \\ 8 & 7 & 1 \\ 4 & 1 & 2 \end{pmatrix}$

(*g*) $\begin{pmatrix} -8 & -5 & 1 \\ -4 & 4 & -5 \\ -12 & -10 & 5 \end{pmatrix}$ (*h*) $\begin{pmatrix} -6 & 11 & -16 \\ 16 & -4 & 8 \\ 6 & -7 & 11 \end{pmatrix}$

(*i*) $\begin{pmatrix} -9 & 4 & 4 \\ 15 & 8 & -12 \\ -3 & 1 & 2 \end{pmatrix}$ (*j*) $\begin{pmatrix} -8 & -5 & 1 \\ -4 & 4 & -5 \\ -12 & -10 & 5 \end{pmatrix}$

$$(k) \begin{pmatrix} -6 & 11 & -16 \\ 16 & -4 & 8 \\ 6 & -7 & 11 \end{pmatrix} \qquad (l) \begin{pmatrix} -9 & 4 & 4 \\ 15 & 8 & -12 \\ -3 & 1 & 2 \end{pmatrix}.$$

2. $(a) \begin{pmatrix} -1 & 2 & 3 \\ 2 & 0 & 1 \\ 1 & 4 & 5 \end{pmatrix} \qquad (b) \begin{pmatrix} -3\alpha & 5\beta & 7\gamma \\ 4\alpha & 2\beta & 3\gamma \\ \alpha & 8\beta & 7\gamma \end{pmatrix}$

6. $\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

11. No, because $AB \neq BA$.

Exercise 1 (*c*)

***4.** (*a*) Yes, Let $A=\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad B=\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$

(*b*) Yes, Let $A=\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B=\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$

(*c*) Yes, Let $A=\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad B=\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$

Exercise 1 (*d*)

1. $\begin{pmatrix} 1 & 3 & -2 \\ -1 & 1 & 7 \\ 2 & 4 & 8 \end{pmatrix}, \quad \begin{pmatrix} 1 & 7 & 0 \\ 3 & 2 & 1 \end{pmatrix},$

$$\begin{pmatrix} -1 & 2 \\ 5 & 3 \\ 6 & 0 \end{pmatrix}.$$

Exercise 1 (*e*)

1. (*a*) -2. (*b*) 24.

(*c*) 23. (*d*) 1.

2. (*a*) x^2+y^2. (*b*) 1.

Exercise 1 (*f*)

1. 0. **2.** 0.

3. 12. **4.** -2.

9. Singular. **10.** Non-singular.

Exercise 1 (*h*)

1. 123.
2. -54.
3. -428.
4. -1.

Exercise 1 (*i*)

1. 9.
2. 7·5.
3. 29.
4. 9.
5. 23·5.
6. 29.
7. a^2.
8. $a^2(p-q)(q-r)(r-p)$.
9. $2a^2 \sin \frac{1}{2}(\beta-\gamma) \sin \frac{1}{2}(\gamma-\alpha) \sin \frac{1}{2}(\alpha-\beta)$.
10. $\frac{1}{2}c^2(q-r)(r-p)(p-q)/(pqr)$.

Exercise 1 (*j*)

1. $x=2, y=3, z=4$.
2. $x=3, y=2, z=1$.
3. $x=2, y=3, z=4$.
4. $x=1, y=2, z=3$.
5. $x=4, y=3, z=2$.
6. $x=1, y=1, z=-1$.
7. $x=1, y=1, z=1$.
8. $x=8/3, y=-1/3, z=0$.
9. $x=3, y=2, z=1$.
10. $x=4, y=5, z=6$.

Exercise 1 (*k*)

3. $\begin{pmatrix} 3 & -4 & -5 \\ -9 & 1 & 4 \\ -5 & 3 & 1 \end{pmatrix}$.

4. $\begin{pmatrix} 8 & -5 & -2 \\ -4 & -3 & 1 \\ -7 & 3 & -1 \end{pmatrix}$.

5. $\begin{pmatrix} -8 & -3 & 1 \\ -1 & 6 & -2 \\ -3 & 1 & -6 \end{pmatrix}$.

6. $\begin{pmatrix} 7 & -11 & -5 \\ 8 & -14 & -5 \\ 6 & -13 & -5 \end{pmatrix}$.

7. $\begin{pmatrix} -9 & 8 & -5 \\ -8 & 7 & -4 \\ -2 & 2 & -1 \end{pmatrix}$.

8. $\frac{1}{9}\begin{pmatrix} 1 & -2 & -1 \\ 5 & 8 & -14 \\ 3 & 3 & -3 \end{pmatrix}$.

9. $\frac{1}{5}\begin{pmatrix} 3 & 2 & -3 \\ -3 & 2 & 2 \\ 2 & -3 & 2 \end{pmatrix}$.

10. $\begin{pmatrix} 2/3 & 59/3 & -9 \\ 1/3 & 40/3 & -6 \\ 0 & -2 & 1 \end{pmatrix}$.

11. $\begin{pmatrix} -1/2 & 1/5 & -8/5 \\ 0 & 1/5 & 2/5 \\ 1/2 & 0 & 1 \end{pmatrix}$.

12. $\frac{1}{25}\begin{pmatrix} 9 & 19 & -4 \\ 4 & 14 & 1 \\ 8 & 3 & 2 \end{pmatrix}$.

13. $\begin{pmatrix} 0 & 1 & -1 \\ 4 & -3 & 4 \\ 3 & -3 & 4 \end{pmatrix}$.

14. $\begin{pmatrix} 4 & 3 & 3 \\ -1 & 0 & -1 \\ -4 & -4 & -3 \end{pmatrix}$.

15. $\begin{pmatrix} 13 & -4 & -2 \\ -2 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$. 16. $\begin{pmatrix} 1 & -2 & 0 \\ -14 & 38 & -9 \\ 0 & -1 & 1 \end{pmatrix}$

17. $-\frac{1}{16}\begin{pmatrix} -8 & -16 & -12 \\ 0 & -4 & -4 \\ -8 & -8 & 4 \end{pmatrix}$. 18. $\begin{pmatrix} 3 & 2 & 6 \\ 1 & 1 & 2 \\ 2 & 2 & 5 \end{pmatrix}$.

19. $\frac{1}{21}\begin{pmatrix} 2 & 3 & -13 \\ -3 & 6 & 9 \\ 5 & -3 & -1 \end{pmatrix}$. 20. $\begin{pmatrix} 2 & 3 & -5 \\ -3 & -5 & 9 \\ 1 & 1 & -2 \end{pmatrix}$.

Exercise 1 (*I*)

1. $x=-2, y=3$.
2. $x=2, y=-3$.
3. $x=1, y=5/2, z=7/2$.
4. $x=1, y=1, z=0$.
5. $x=1/3, y=2/3, z=2$.
6. $x=1, y=1, z=2$.
7. $x=2, y=-1, z=1$.
8. $x=-3, y=2, z=-1$.
9. Consistent.
10. Inconsistent.
11. Consistent.
12. Inconsistent.
13. $x=2k-1, y=-2k+3, z=k$.
14. $x=\frac{1}{8}(23-7k),\ y=\frac{1}{8}(1-k), z=k$.
15. $x=y=z=w=1$.
16. $x=\frac{5}{3}-\frac{3}{2}k, y=-\frac{5}{2}+\frac{1}{2}k, z=k$.
17. $x=y=z=0$.
18. $x=y=z=0$.
19. $x=k, y=-2k, z=k$.
20. $x=3k_1-2k_2, y=k_1, z=k_2$.

Test Your Understanding I

1. (*d*) 2. (*b*). 3. (*c*). 4. (*d*). 5. (*b*).
6. (*a*). 7. (*b*). 8. (*a*). 9. (*b*). 10. (*c*).

Review Exercise I

1. $\begin{pmatrix} -21 & -2 & 15 \\ 13 & -4 & 17 \end{pmatrix}$. 2. $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

3. $\begin{pmatrix} -1 & 1 & -1 \\ 3 & -3 & 3 \\ 5 & -5 & 5 \end{pmatrix}$. 4. $-\frac{1}{13}\begin{pmatrix} 1 & -8 \\ -2 & 3 \end{pmatrix}$.

5. $\begin{pmatrix} 1/a & 0 \\ 0 & 1/b \end{pmatrix}$. 6. $\begin{pmatrix} 25 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 25 \end{pmatrix}$

7. $\frac{1}{21}\begin{pmatrix} 2 & 3 & -13 \\ -3 & 6 & 9 \\ 5 & -3 & -1 \end{pmatrix}$ **9.** $x=2, y=1.$

10. $\frac{1}{16}\begin{pmatrix} 0 & -20 \\ 4 & 17 \end{pmatrix}.$ **11.** $x=1, y=3, z=5.$

12. $\frac{1}{35}\begin{pmatrix} 13 & 12 & 5 \\ 5 & 10 & 10 \\ 12 & 3 & 10 \end{pmatrix}$; $x=y=z=1.$

13. $\begin{pmatrix} 1 & -p & p^2 \\ 0 & 1 & -p \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ -q & 1 & 0 \\ q^2 & -q & 1 \end{pmatrix}.$

$$\begin{pmatrix} 1 & -p & p^2 \\ -q & pq+1 & -p(pq+1) \\ q^2 & -p(pq+1) & p^2q^2+pq+1 \end{pmatrix}.$$

14. $\begin{pmatrix} 2 & 2 & 1 \\ -2 & 1 & 2 \\ 1 & -2 & 2 \end{pmatrix}, \quad \begin{pmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$

$$\frac{1}{9}\begin{pmatrix} 2 & 2 & 1 \\ -2 & 1 & 2 \\ 1 & -2 & 2 \end{pmatrix}, \quad \frac{1}{9}\begin{pmatrix} 1 & 2 & 8 \\ -4 & 7 & 4 \\ 8 & -4 & 1 \end{pmatrix}.$$

Exercise 2 (*a*)

1. (*a*) $\mathbf{R}\sim\{0, 2\}$, (*b*) $]-\infty, 0[\ \cup\ ]3, \infty[$.
2. (*a*) $\mathbf{R}^+ \cup \{0\}$, (*b*) $\mathbf{R}$.
3. (*a*) $\mathbf{R}$, (*b*) $[1, \infty[$.
4. $\mathbf{R} \sim \{-6/5\}$, $\mathbf{R} \sim \{0\}$.
5. $\mathbf{R} \sim \{-4/3\}$; $\mathbf{R} \sim \{2/3\}, f^{-1}(x)=\frac{3-4x}{3x-2}$.

Exercise 2 (*c*)

1. 11. **2.** $\frac{5}{3}$. **3.** 10. **4.** $\frac{1}{3}$. **5.** 8.
6. 30. **7.** 0. **8.** $-\frac{1}{3}$. **9.** $-\frac{1}{36}$. **10.** $-\frac{1}{54}$.
11. $\frac{1}{4}$. **12.** $\frac{1}{2}$. **13.** -1. **14.** $-\frac{3}{4}$. **15.** -12.

Exercise 2 (*d*)

1. 2. **2.** -1. **3.** 0. **4.** 1. **5.** 0.
6. 0. **7.** 0. **8.** 2. **9.** 1. **10.** 625.

Exercise 2 (*e*)

1. $\frac{2}{3}$. **2.** $\frac{5}{3}$. **3.** $\frac{4}{5}$ **4.** $\frac{3}{2}$. **5.** $\frac{5}{4}$.

6. $\frac{3}{4}$. **7.** 2. **8.** 0. **9.** 4. **10.** $\frac{1}{2}$.

11. 0. **12.** 2. **13.** 1.

Exercise 2 (*f*)

1. 1. **2.** 1. **3.** e^{-1}. **4.** e^3. **5.** $\frac{2}{3}$.

6. $\frac{4}{5}$. **7.** 0. **8.** 0. **9.** $\frac{1}{4}$. **10.** $\frac{5}{3}$.

Exercise 2 (*g*)

1. $+\infty$. **2.** $-\infty$. **3.** $+\infty$. **4.** $+\infty$. **5.** 0.

6. 0. **7.** $-\infty$. **8.** $-\infty$. **9.** $-\infty$. **10.** $+\infty$.

Exercise 2 (*h*)

3. Discontinuous at $x=1$ and continuous everywhere else.

4. $\frac{2}{5}$.

5. (*i*) continuous (*ii*) discontinuous (*iii*) discontinuous (*iv*) discontinuous (*v*) continuous.

6. (*i*) Discontinuous (*ii*) continuous (*iii*) discontinuous.

8. Discontinuous at $0, \pm 1, \pm 2, \pm 3, \ldots$

9. No, consider $f(x)=-1$ if $x < 0, f(x)=1$ if $x \geqslant 0$ at point $x=0$.

10. Discontinuous at the points $n+\frac{1}{2}$, where n is any integer.

Test Your Understanding II

1. (*d*). **2.** (*b*). **3.** (*b*). **4.** (*c*) **5.** (*a*).

6. (*d*). **7.** (*c*). **8.** (*d*). **9.** (*b*) **10.** (*d*).

Review Exercise II

1. $\mathbf{R} \sim \{-1, 1\}$.

2. (*a*) $]-\infty, -1[\cup]1, \infty[$, (*b*) $\{x : |x| \leqslant \sqrt{2}\}$, (*c*) $\{x : x \geqslant 2\}$.

3. $]-\infty, -1[\cup [0, \infty[$.

4. (*a*) $\{x : x \geqslant 0\}$, (*b*) $\{x : |x| \leqslant 4\}$, (*c*) $[0, \infty[$.

5. (*a*) Domain$=\mathbf{R}$, Range$=[-2, 2]$.

(*b*) Domain$=\mathbf{R}$, Range$=[-3, 3]$

(*c*) Domain$=\mathbf{R} \sim \{n\pi, : n \in \mathbf{Z}\}$, Range$=]-\infty, \infty[$,

(*d*) Domain$=\{x : |x| \geqslant \frac{1}{2}\}$, Range$=[-\pi/4, \pi/4]$,

(*e*) Domain$=\mathbf{R}$, Range$=]-3\pi, 3\pi[$,

(f) Domain $= \left\{x : |x| \geqslant \frac{1}{4}\right\}$,

Range $=[-5\pi, 0]\sim\left\{-\frac{5}{2}\pi\right\}$.

6. (a) Domain $=\mathbf{R}$, Range $=]0, \infty[$,
   (b) Domain $=\mathbf{R}$, Range $=]0, \infty[$,
   (c) Domain $=\mathbf{R}$, Range $=[5, \infty[$,
   (d) Domain $=\{x : x > 0\}$, Range $=\mathbf{R}$,
   (e) Domain $=\{x : x < 1\}$, Range $=\mathbf{R}$,
   (f) Domain $=\{x : x > -\frac{1}{2}\}$, Range $=\mathbf{R}$.
7. ϕ ; it does not define a function.
8. $\frac{1}{2}$. 9. $\frac{1}{3}$. 10. 0.
11. $\lim\limits_{-1-} f(x)=-1$, $\lim\limits_{-1+} f(x)=1$, $\lim\limits_{x\to 1} f(x)=1$.
12. $[0, 3/\sqrt{2}]$. 13. $\frac{1}{5}$. 14. $-\frac{4}{9}\sqrt{3}$.
15. $\frac{1}{2}$. 16 $\frac{1}{2}$. 17. 1.
18. $\sqrt{2}$. 19. $2/\pi$.
21. $g(x)=\begin{cases} 2+x, & 0 \leqslant x \leqslant 1 \\ 2-x, & 1 < x < 2 \\ 4-x, & 2 < x \leqslant 3 \end{cases}$ Discontinuous at $x=1$ and $x=2$.
22. Domain $=\mathbf{R}$, Range $=[0, 1[$. The function is not one-to-one.
23. $[-2, 1[\sim\{0\}$. 24. $\ln 4$.
25. $a=-\frac{3}{2}$, $b\neq 0$, $c=\frac{1}{2}$. 26. $a+b$.
27. $f(0+)=0$, $f(0-)=-1$, Discontinuous at $x=0$.

Exercise 3 (a)

1. 2π. 2. 8π.
3. 4 per sec. 4. 80.
5. The entries in the third column are 1, 29, -2, 12 and 0.
6.

$f(x)$	Average rate of change	$f(5)$
$x-9$	1	1
$2x^2+x+1$	$2h+21$	21
x^2+5	$h+10$	10
x^3	$h^2+15h+75$	75
6	0	0

Exercise 3 (*b*)

1. 0. **2.** 1. **3.** 2. **4.** 2.
5. a. **6.** $2x$. **7.** $2x$. **8.** $4x$.
9. $-2x$. **10.** -1 when $x < -1$, and 1 when $x > -1$.
11. -1 when $x < 0$, and 1 when $x > 0$.
12. 1 when $x < 0$, and -1 when $x > 0$.
13. -2 when $x < 0$, and 2 when $x > 0$.
14. Those in problems 1 to 9.
15. -1 for the function in problem 10 and 0 in problems 11—13.
16. (*a*) $5x^4$, (*b*) $10x^9$, (*c*) $100x^{99}$,
(*d*) $1000x^{999}$, (*e*) $\frac{1}{2}(x+2)^{-1/2}$, (*f*) $-\frac{1}{2}\sqrt{5}x^{-1/2}$,
(*g*) $\frac{3}{2}(3x+7)^{-1/2}$.

Exercise 3 (*c*)

2. (*a*) $6x+4$, (*b*) $9x^2-10x$, (*c*) $6x(4x^2+1)$,
(*d*) $45x^8+42x^6+1$, (*e*) $15x^{14}-11$, (*f*) $100^2\, x^{99}$,
(*g*) $a_1+2a_2\, x+3a_3\, x^2+\cdots\cdots+na_n\, x^{n-1}$.
3. (*a*) $-3x^{-4}+1+3x^2$, (*b*) $-x^{-8}(7+5x^2)$,
(*c*) $-3(x^{-2}-2x^5)$, (*d*) $-4x^{-13}(15+x^8)+1$,
(*e*) $2x^{-2}(x+x^{-1})(x^2-1)$, (*f*) $-6x^{-3}(x^{-2}+2)^2$.
4. (*a*) $-(5x+7)^{-2}$, (*b*) $(15x^2+100x+27)(3x+10)^{-2}$,
(*c*) $x(x^3-3x-2)(x^2-1)^{-2}$, (*d*) $\dfrac{-2(4x^4+7x^2+2)}{x^5(1+2x^2)^2}$,
(*e*) $\dfrac{-x^2+2x+1}{(x^2+1)^2}$, (*f*) $\dfrac{18x^{20}+405x^{12}-44x^8-30}{x^4(2x^8+5)^2}$,
(*g*) $\dfrac{2cpx+(bp+3cq)+(2bq+4cr)x^{-1}+(3br+aq)x^{-2}+2arx^{-3}}{(p+qx^{-1}+rx^{-2})^2}$
5. (*a*) $8x(x^2+1)\{(x^2+1)^2+3\}$; 2240.
(*b*) $12x^2(x^3+1)\{(x^3+1)^2+5\}$; 37152.
(*c*) $12x(x^2+1)\{(x^2+1)^2-5\}^2$; 48000.
(*d*) $54x^2(2x^3-1)^2\{(2x^3-1)^3-3\}^2$; $54(101160)^2$.

Exercise 3 (*d*)

1. $14(2x+11)^6$. **2.** $156(3x-7)^{51}$.
3. $14(7x+6)$. **4.** $-106(-2x+1)^{52}$.
5. $\dfrac{-6}{(1+3x+3x^2+x^3)}$. **6.** $\dfrac{-4(1+3x+9x^2)}{(2x+3x^2+6x^3)^3}$.
7. $3x^2+86x+519$. **8.** $-7(x+5)^{-8}+99(9x+1)^{10}$.

9. $\dfrac{-(2x+3)(14x+27)}{(7x+9)^4}$.

10. $\dfrac{(84x+414x^2-220x^3-590x^4-546x^5)(3+5x^2+7x^3)^8}{(9+11x^2+13x^3)^{12}}$

11. $4(2x+1)(x^2+x+1)\{(x^2+x+1)^2+3\}$.

12. $10\{18x^2(2x^3-3)^2+1\}\{(2x^3-3)^3+x\}^9$.

13. $2\{x+(x+1)^2+(x^2+x+1)^2\}\{1+2(x+1)+2(x^2+x+1)(2x+1)\}$.

14. $7\{(2x+5)^3+(9x-5)^{-3}\}^6\ \{6(2x+5)^2-27(9x-5)^{-4}\}$.

15. $16[\{(1+x^2)^2+x^4\}^3+x^8]^{15}[3\{(1+x^2)^2+x^4\}^2\{4x(1+x^2)+4x^3\}+8x^7]$.

Exercise 3 (*e*)

1. $\dfrac{2}{3}(2x+3)^{2/3}$.

2. $\dfrac{15}{7}(3x-5)^{-2/7}$.

3. $4x(3x^2+1)^{-1/3}$.

4. $\dfrac{1}{6}(4x-5)(2x^2-5x+3)^{-5/6}$.

5. $\dfrac{2}{5}(2x+3)^{-4/5}+\dfrac{1}{3}(x-2)^{-2/3}$.

6. $\dfrac{2}{3}x^{-1/3}(x^2+1)^{1/2}+x^{5/3}(x^2+1)^{-1/2}$.

7. $\dfrac{3}{4}x^{-1/4}(2x-1)^{1/2}+x^{3/4}(2x-1)^{-1/2}$.

8. $\dfrac{3}{2}(3x-2)^{-1/2}(x+4)^{1/3}+\dfrac{1}{3}(3x-2)^{1/2}(x+4)^{-2/3}$.

9. $1+\dfrac{5}{3}x^{-1/6}+\dfrac{2}{3}x^{-1/3}$.

10. $(x^{2/3}-x^{3/4})^{-1/2}\left(\dfrac{1}{3}x^{-1/3}-\dfrac{3}{8}x^{-1/4}\right)$.

11. $2(4x+1)^{-1/2}+\dfrac{4}{3}x(x^2-3)^{-1/3}$.

12. $2x(2x^2+1)^{-1/2}-(4-x)^{-2/3}$.

Exercise 3 (*f*)

1. $3\cos(3x+4)$.

2. $-2\cos x\sin x$.

3. $\tan(4-x^2)-2x^2\sec^2(4-x^2)$.

4. $-6\cot^2(2x-1)\csc^2(2x-1)$.

5. $(4x+1)\sec(2x^2+x+1)\tan(2x^2+x+1)$.

6. $2x\csc(3x-5)-3x^2\csc(3x-5)\cot(3x-5)$.

7. $3\sin^2 x\cos^3 x-2\cos x\sin^4 x$.

8. $\cos(\tan x)\sec^2 x$.

9. $-\dfrac{1}{2}(\cos x)^{-1/2}\sin x$.

10. $\dfrac{-1}{2\sqrt{x}}\sin\sqrt{x}$.

11. $\dfrac{\sin x-\cos x}{(\sin x+\cos x)^2}$.

12. $6x^2\tan(x^3)\sec^2(x^3)$.

13. $\cos^2(2x-1)-4x\cos(2x-1)\sin(2x-1)$.

14. $-\csc^2(3x)(\cot 3x)^{-2/3}$.

15. $-2x\csc x^2\cot x^2-2\csc^2 x\cot x$.

Exercise 3 (*g*)

1. $\dfrac{|a|}{a\sqrt{(a^2-x^2)}}$.
2. $\dfrac{2x}{1+x^4}$.
3. $\cos^{-1} 2x-\dfrac{2x}{\sqrt{1-4x^2}}$
4. $\dfrac{-3}{|x|\sqrt{x^6-1}}$
5. $\dfrac{1}{2x^2+6x+5}$
6. $\dfrac{9(\sin^{-1} 3x)^2}{\sqrt{1-9x^2}}$.
7. $\dfrac{1}{2\sqrt{x}}\cos^{-1}\sqrt{x}-\dfrac{1}{2\sqrt{1-x}}$.
8. $\dfrac{2}{1+x^2}$
9. $\csc^{-1}\dfrac{x}{3}-\dfrac{3x}{|x|\sqrt{x^2-9}}$
10. $\dfrac{2\sec^{-1}(x+1)}{|x+1|\sqrt{x(x+2)}}$.

Exercise 3 (*h*)

1. $\dfrac{2}{1+x^2}$.
2. $\dfrac{2}{1+x^2}$
3. $-\dfrac{2}{\sqrt{1-x^2}}$.
4. $-\dfrac{3}{\sqrt{1-x^2}}$.
5. $\dfrac{3}{\sqrt{1-x^2}}$.
6. $\dfrac{1}{2\sqrt{x}(1+x)}$.
7. $\dfrac{3}{1+x^2}$.
8. $-\dfrac{1}{2}$.
9. $\dfrac{2a}{\sqrt{1-a^2x^2}}$
10. $\dfrac{1}{3}\dfrac{1}{x^{2/3}(1+x^{2/3})}$.
11. $\dfrac{1}{2(1+x^2)}$.
12. $-\dfrac{1}{2}$.
13. 1.
14. $-(1+x^2)$.
15. $-\dfrac{1}{2}$.

Exercise 3 (*i*)

1. $2e^{2x+3}$.
2. $(4x+5)e^{2x^2+5x-7}$.
3. $e^{\sin x}\cos x$.
4. $\dfrac{-e^{\cos^{-1}\sqrt{x}}}{2\sqrt{x-x^2}}$
5. $\dfrac{e^{2x}}{\sqrt{e^{2x}+1}}$.
6. $\dfrac{e^{\sqrt{x}}}{2\sqrt{x}}\sec^2(e^{\sqrt{x}})$

7. $\dfrac{e^x}{(e^{2x}+1)^{3/2}}$ 8. $\tan^{-1}(e^{x^2})+\dfrac{2x^2\,e^{x^2}}{1+e^{2x^2}}$

9. $\dfrac{1}{2}\,x\,(4+\sqrt{x}\,)\,e^{\sqrt{x}}$ 10. $-\dfrac{1}{4}\,e^x\,\{x(4-e^{2\sqrt{x}}\,)\}^{-1/2}$.

Exercise 3 (*j*)

1. $\dfrac{1}{2x}$. 2. $-2\tan x$.

3. $\dfrac{4x}{1-x^4}$. 4. $\dfrac{x}{x^2+4}$.

5. $\dfrac{e^x}{2(e^x+1)}$. 6. $\left(\dfrac{1}{x}+2x\ln x\right)e^{x^2}$.

7. $\dfrac{4x\ln(x^2+1)}{x^2+1}$. 8. $-\dfrac{1}{x(\ln x)^2}$.

Exercise 3 (*k*)

1. $2x\,.\,10^{x^2}\ln 10$. 2. $5^{\sin x}\cos x\ln 5$.

3. $-2x\sin x^2\ln 2.2^{\cos x^2}$. 4. $\dfrac{2}{2x+1}\log_5 e$.

5. $\log_5\tan^2 x+4x\csc 2x\log_5 e$.

6. $\cos x\log_{10}(x^2+1)+\dfrac{2x\sin x}{x^2+1}\log_{10} e$.

7. $2\sqrt{3}x(x^2+4)^{\sqrt{3}-1}$. 8. $\dfrac{1}{2\sqrt{x}}\,(e+1)\,(\sqrt{x}+2)^e$.

Exercise 3 (*l*)

1. $(x-1)(3x-5)\sin x+(x-1)^2(x-2)\cos x$.

2. $\cos x\cos^{-1}x\ln x+\dfrac{1}{x}\sin x\cos^{-1}x-\dfrac{\sin x\ln x}{1-x^2}$.

3. $\dfrac{12-37x-14x^2}{12x^{1/2}(1-2x)^{2/3}(x+2)^{5/4}}$.

4. $-\dfrac{3x^2+12x+11}{(x+1)^2(x+2)^2(x+3)^2}$.

5. $x^{(x+x^x)}\left\{\left(\dfrac{1}{x}\right)+(\ln x)\ln(ex)\right\}$.

6. $x(x^x)^x\ln(ex^2)$.

7. $x^{\sin^{-1}x}\left\{\left(\dfrac{1}{x}\right)\sin^{-1}x+(\ln x)/\sqrt{(1-x^2)}\right\}$.

8. $(\sin x)^{\cos^{-1}x}\left\{\cos^{-1}x\cot x-\dfrac{\ln\sin x}{\sqrt{(1-x^2)}}\right\}$.

9. $(\tan x)^{x}(\ln\tan x+2x\csc 2x)+x^{\tan x}(\sec^2 x\ln x+(1/x)\tan x)$.

10. $(\tan x)^{\cot x}\csc^2 x\ln(e\cot x)-(\cot x)^{\tan x}\sec^2 x\ln(e\tan x)$.

11. $(\sin x)^{\cos x}(\cot x\cos x-\sin x\ln x)$
$\quad+(\cos x)^{\sin x}(-\tan x\sin x+\cos x\ln\cos x)$.

12. $x^{x}\ln(ex)+(\tan x)^{\ln x}\left\{\dfrac{\ln\tan x}{x}+\dfrac{\ln x}{\sin x\cos x}\right\}$.

Exercise 3 (*m*)

1. $-\dfrac{ax+hy}{hx+by}$.

2. $-\dfrac{ax+hy+g}{hx+by+f}$.

3. $\dfrac{y^2}{x(1-y\ln x)}$.

4. $\dfrac{y(x-y)}{x(x+y)}$.

5. $\dfrac{y\{(a+bx)y-bx^2\}}{x(y-x)(a+bx)}$.

6. $\dfrac{y}{x}$.

7. $\dfrac{\cos y+y\sin x}{\cos x+x\sin y}$.

8. $-\dfrac{5x^4\sin y+y^5\cos x}{x^5\cos y+5y^4\sin x}$.

9. $\dfrac{\ln\sin y+y\tan x}{\ln\cos x-x\cot y}$.

10. $-\dfrac{(yx^{y-1}+y^x\ln y)}{x^y\ln x+xy^{x-1}}$.

Exercise 3 (*n*)

1. $-\cot\theta$.

2. $-\dfrac{b}{a}\cot\theta$.

3. $\dfrac{b}{a}\csc\theta$.

4. $-\dfrac{b}{a}\tan\theta$.

5. $\tan 2\theta$.

6. $\tan\theta$.

7. $\tan\left(\dfrac{\theta}{2}\right)$.

8. $\dfrac{1}{t}$.

9. $-\tan 3t$.

10. 0.

Exercise 3 (*o*)

1. $48(2x+3)^2$.

2. $16e^{4x}$.

3. $\dfrac{50}{(5x+4)^3}$.

4. $\dfrac{-9}{(4-3x)^2}$

5. $-\dfrac{a^2}{(ax+b)^2}+\dfrac{c^2}{(cx+d)^2}$.

6. $\dfrac{x}{(1-x^2)^{3/2}}$

7. $-2\cos 2x+18\cos 6x$.

8. $2(\sec^2 x\tan x+\csc^2 x\cot x)$

9. $(a^2+b^2)e^{ax}\sin\left(bx+2\tan^{-1}\left(\dfrac{b}{a}\right)\right)$.

Exercise 3 (*p*)

1. $\dfrac{1}{2}(n+2)!\,3^n(3x+4)^2$.

2. $2^n\,.\,n!$

3. $\dfrac{(-1)^n \cdot n!\, 2^n}{(2x+7)^{n+2}}$.

4. $\dfrac{n!\, b^n}{(a-bx)^{n+1}}$.

5. $\dfrac{(-1)^n (n+1)!\, 3^n}{(3x+8)^{n+2}}$.

6. $\dfrac{11(-1)^n n!\, 3^{n-1}}{(3x+4)^{n+1}}$.

7. $(-1)^{n-1} 2^{-n} \cdot 1.3.5 \ldots\ldots (2n-3) \cdot 9^n (9x+8)^{(1-2n)/2}$.

8. $\dfrac{(-1)^{n-1} (n-1)!}{(x+2)^n}$.

9. $5^x (\ln 5)^n$.

10. $3.9^n \cdot e^{9x}$.

11. $3^n \sin\left(3x+4+\dfrac{1}{2} n\pi\right)$.

12. $\cos\left(x+2+\dfrac{1}{2} n\pi\right)$.

13. $\sqrt{3}\, e^{3x} \sin\left(2x+n \tan^{-1}\left(\dfrac{2}{3}\right)\right)$.

14. $e^x \cos\left(x+5+\dfrac{1}{4} n\pi\right)$.

15. $\dfrac{1}{2}\left[\cos\left(2x+\dfrac{1}{2} n\pi\right)-\cos\left(4x+\dfrac{1}{2} n\pi\right)\right]$.

16. $\dfrac{1}{16}\left[2\cos\left(x+\dfrac{1}{2} n\pi\right)-3^n \cos\left(3x+\dfrac{1}{2} n\pi\right)\right.$

$\left.-5^n \cos\left(5x+\dfrac{1}{2} n\pi\right)\right]$.

17. $2^{n-1} \cos\left(2x+\dfrac{1}{2} n\pi\right)+\dfrac{1}{2} 4^{n-1} \cos\left(4x+\dfrac{1}{2} n\pi\right)$.

18. $\dfrac{1}{4}\left[2^n \cos\left(2x+\dfrac{1}{2} n\pi\right)+4^n \cos\left(4x+\dfrac{1}{2} n\pi\right)\right.$

$\left.+6^n \cos\left(6x+\dfrac{1}{2} n\pi\right)\right]$.

19. $\dfrac{1}{2}\left[[(\sqrt{37})^n e^x \cos(6x+n \tan^{-1} 6)\right.$

$\left.+(\sqrt{5})^n e^x \cos(2x+n \tan^{-1} 2)\right]$.

20. $\dfrac{1}{4}\,(\sqrt{13})^n e^{3x} \sin\left(2x+n \tan^{-1}\left(\dfrac{2}{3}\right)\right)$

$+5^n e^{3x} \sin\left(4x+n \tan^{-1} \dfrac{4}{3}\right)+(\sqrt{39})^n \sin(6x+n \tan^{-1} 2)\Big]$.

Test Your Understanding III

1. (*c*) 2. (*c*) 3. (*d*) 4. (*d*) 5. (*b*)
6. (*a*) 7. (*a*) 8. (*c*) 9. (*b*) 10. (*b*).

Review Exercise III

1. (a) $2ax-\frac{b}{x^2}$.

2. (a) $3\cos 3x$ (b) $2\cos(2x+3)$.

3. (a) $-2\sin 2x$ (b) $-\sin 2x$ (c) $\frac{-\sin x}{2\sqrt{\cos x}}$.

4. (a) $2\sec^2 2x$ (b) $3\sec^2(3x+1)$ (c) $-\csc^2 x$.

5. (a) $-\frac{1}{1+x^2}$ (b) $\frac{1}{\sqrt{(1-x^2)}}$ (c) $\frac{1}{1+x^2}$.

6. $2x\, e^{x^2}$.

7. (a) $\frac{(8x^n+10x+3)(2x+3)}{2(x+1)^{3/2}\sqrt{x}}$ (b) $\frac{(11x^2+2x-18)(x+4)^{1/2}}{3\sqrt{x}(4x-3)^{7/3}}$.

8. (a) $\frac{-2\sin x}{(1+\cos x)^2}$ (b) $e^{m\sin^{-1}x}\left(\frac{m\sin nx}{\sqrt{1-x^2}}+n\cos nx\right)$.

9. (a) $e^x\left[\ln(1+x^2)+\frac{2x}{1+x^3}\right]$

(b) $\frac{1}{3}x\cot\left(\frac{1}{3}x^2-1\right)\Big/\left[\sqrt{\ln\left\{\sin\left(\frac{1}{3}x^2-1\right)\right\}}\right]$.

10. (a) $x^x \ln(ex)$. (a) $(\sin x)^x\,[x\cot x+\ln\sin x]$.

11. (a) $(\sin x)^{\ln x}\left[\cot x\,(\ln x)+\frac{1}{x}\ln(\sin x)\right]$.

(b) $x^{\sin x}\left[\frac{1}{x}\sin x+\cos x\ln x\right]+(\sin x)^x\,[\ln(\sin x)+\cot x]$.

12. (a) $\frac{y(x\ln y-y)}{x(x-x\ln y)}$.

(b) $\frac{y^{\cot x}\ln y\cos^2 x\,(1+x^2)-y\,(\tan x)^{y-1}}{(1+x^2)\,[\cot x\, y^{\cot x-1}+\ln\tan^{-1}x(\tan^{-1}x)^y]}$.

13. (a) $\frac{2}{1+x^2}$. (b) $-\frac{2}{\sqrt{(1-x^2)}}$.

14. (a) -1. (b) $\frac{-4\sin^2 x}{\cos^4 x+\sin^4 x}$

15. (a) $\frac{1}{2}$. (b) $\frac{1}{2}$. (c) $-\frac{1}{2\sqrt{x}(1+x)}$

Exercise 4 (a)

1. (a) 8 m/sec, 4 m/sec², (b) 11·5 m/sec, 1·5 m/sec².
(c) 0, −8 m/sec².

2. A=1, B=−1, C=6 ; 22 m.

Exercise 4 (*b*)

1. 29·4 m.
2. 28 m/sec.
3. 5·6 m/sec vertically downwards ; 10 m.
4. $y=x\sqrt{3}-\frac{2}{5}x^2$; $\frac{5\sqrt{3}}{14}$ sec
5. 28 m/sec.
6. 7 m/sec.
7. 30 m.

Exercise 4 (*c*)

1. $\frac{25}{3\pi}$ cm/sec.
2. 6 cm^3/sec.
3. 250 π cm^3/min.
4. 432 km/hr.
5. ·05 cm^3/sec.
6. 2m/sec.
7. $\frac{1}{15\pi}$ cm/sec.
8. 1·5 m/sec away from O.
9. $\frac{4}{9\pi}$ cm/m.

Exercise 4 (*d*)

1. $3(x^2-1)\,dx$.
2. $\frac{-x^2+4x+4}{(x^2+4)^2}\,dx$.
3. $5(2x+1)^{3/2}\,dx$.
4. $\frac{1}{2}(5\cos 5x+\cos x)\,dx$
5. $e^{2x}\left[\frac{1}{x-1}+2\ln(x-1)\right]dx$.
6. $-e^{-3x}(3\cos 2x+2\sin 2x)\,dx$.
7. ·06.
8. ·42.
9. ·02.
10. ·03.
11. 7·9325.
12. 5·0133.
13. 9·99.
14. 2·0007.
15. 1009 cm^3 ; 603·6 cm^2.
16. 290·88 π cm^3 ; 144·96 π cm^2.
17. 2 per cent.
18. ·1 cm.
19. ·6 per cent.
20. $\pm$ 30 cm^3.

Exercise 4 (*e*)

1. (*i*) $x(x'+g)+y(y'+f)+gx'+fy'+c=0$,
   (*ii*) $xx'/a^2+yy'/b^2=1$,
   (*iii*) $xx'/a^2-yy'/b^2=1$,
   (*vi*) $xy'+x'y=2c^2$,
   (*v*) $yy'=2a(x+x')$.
2. (*i*) $x\cos\theta/a+y\sin\theta/b=1$,
   (*ii*) $x\sec\theta/a-y\tan\theta/b=1$,
   (*iii*) $x\cos\theta+y\sin\theta=a\sin\theta\cos\theta$,
   (*iv*) $x\cos\theta/2-y\sin\theta/2=a\theta\cos\theta/2-2a\sin\theta/2$.

3. $x-4y+3=0$.
4. $((6+\sqrt{3})/3, -2\sqrt{3}/9), ((6-\sqrt{3})/3, 2\sqrt{3}/9)$.
5. $y=12x+16, y=12x-16$.
6. $a^2\cos^2\alpha+b^2\sin^2\alpha=p^2$.

Exercise 4 (*f*)

1. (*i*) $xy'-x'y=0$, (*ii*) $xy'+2ay=(x'+2a)y'$, (*iii*) $a^2x/x'-b^2y/y'=a^2-b^2$, (*iv*) $a^2x/x'+b^2y/y'=a^2+b^2$.
2. (*i*) $ax\sec\theta-by\cot\theta=a^2-b^2$,
   (*ii*) $ax\cos\theta+by\cot\theta=a^2+b^2$,
   (*iii*) $x\cos(\theta/2)+y\sin(\theta/2)=a\theta\cos(\theta/2)+2a\sin(\theta/2)$,
   (*iv*) $x\sin\theta-y\cos\theta=a\sin\theta\cos\theta\cos 2\theta$.
3. $y+20x-140=0$.
4. $3x-2y=2$.
5. $y=4x-144$.
6. $27A^2=B(3A^3+B^2)^2$.

Exercise 4 (*g*)

3. $\frac{1}{3}(\sqrt{37}-4)$.

7. (*a*), (*b*), (*c*). Neither the hypotheses are valid nor the conclusion is valid. (*d*) Both the hypotheses and the conclusion are valid.

12. $(\sqrt{6}, \sqrt{2})$.

13. (*a*) $\frac{5}{4}$, (*b*) $\frac{1}{5}\sqrt{(91/3)}$,

    (*c*) $1+(\cdot 1)(\ln 1\cdot 1)^{-1}$, (*d*) $1+\left[e\ln\left(1+\frac{1}{e}\right)\right]^{-1}$

14. (*a*), (*b*), (*d*). Neither the hypotheses nor the conclusion are valid. (*c*) Hypotheses are not valid but the conclusion is valid.

Exercise 4 (*j*)

1. min. at $\frac{1}{3}(6+\sqrt{3})$ and max. at $\frac{1}{3}(6-\sqrt{3})$

   $2/3\sqrt{3}$ is a maximum value and $-2/3\sqrt{3}$ is a minimum value.
2. Maximum at $x=1$, minimum at $x=6$.
3. Maximum at $x=\pi/3$, minimum at $x=5\pi/3$.
4. Maximum at $x=\pi/6$, minimum at $x=-\pi/6$.
5. Maximum at $x=\pi/6$, minimum at $x=5\pi/6$.
6. Maximum at $x=\pi/6$.
7. 5 is a maximum value and 50 is a minimum value.
8. Maximum at $x=1$ and minimum at $x=2$.
9. 12, 12.
10. 16, 16.
11. 64, 96.
12. Each side $\sqrt{96}$ cm ; perimeter $16\sqrt{6}$ cm.

13. Each side 10 cm.

14. Diameter$=40/(\pi+4)$ m, width of rectangle$=20/(\pi+4)$ m.

16. Radius$=\sqrt{(200/3)}$, height$=\sqrt{(800/3)}$.

17. 72π cm.

Test Your Understanding IV

1. (*c*).	**2.** (*a*).	**3.** (*c*).	**4.** (*b*).	**5.** (*a*).
6. (*d*).	**7.** (*d*).	**8.** (*c*).	**9.** (*d*).	**10.** (*d*).

Review Exercise IV

1. $t>\frac{1}{3}$

2. 4 m/min.

3. 28 m/sec.

4. 24 m/sec^2.

5. ·75 m/sec.

6. 3·0123.

7. (*a*) $\pm 2\pi$ cm^2, (*b*) $\pm 5\pi$ cm^3, (*c*) $\pm$ 2 per cent, (*d*) $\pm$ 3 per cent.

8. 14 m/sec ; -24 m/sec^2.

9. In the ratio $\pi : 4$.

10. At a distance $160/\sqrt{3}$ km from D on the line CD where D is the mid-point of AB.

11. Maximum at -3, 1 and minimum at 0.

12. 4000 cm^3.

13. Height$=(2/\sqrt{3})r$; radius of the base$=\sqrt{2/3}\,r$.

14. $\sqrt{(a^2+b^2)}$.

15. If $a>b$, max. value is a and min. value is b ; if $a<b$, max. value is b and min. value is a.

16. $(a+b)^2$.

18. $x+3y=\pm 4,\ x+3y=\pm 8$.

Exercise 5 (*a*)

1. $\frac{1}{5}x^5+C$.

2. $\frac{2}{3}x^{3/2}+C$.

3. $\sin x+C$.

4. $\sec x+C$.

5. $\frac{3}{5}x^5+C$.

6. $\frac{1}{2}x^2+ln\,|\,x\,|\,+C$.

7. $\frac{1}{3}x^3+x^2+x+C$.

8. $e^x-\cos x+C$.

9. $2\,ln\,|\,x\,|\,+\frac{2}{3}x^{3/2}+5\tan x+C$.

10. $\frac{1}{6}x^6-\frac{4}{5}x^5+5x+C$.

11. $4\,ln\,|\,x\,|\,-\frac{3}{x}-\frac{1}{x^2}+C$.

Exercise 5 (*b*)

1. $\frac{1}{12}x^{12}+C$.

2. $\frac{7}{4}x^{4/7}+C$.

3. $2\sqrt{x}+C.$ 4. $-\frac{1}{4x^4}+C.$

5. $\frac{1}{5}x^5-x^2+5x+C.$ 6. $\frac{1}{4}x^4+\frac{1}{2x^2}+C.$

7. $\frac{4}{3}x^3+6x^2+9x+C.$ 8. $\frac{1}{3}x^3+\frac{3}{2}x^2+2x+C.$

9. $\frac{10}{3}x^3-\frac{7}{2}x^2-12x+C.$ 10. $\frac{3}{17}x^{17/3}+\frac{6}{7}x^{7/6}+3x^{3/2}+C.$

11. $ln\mid x-5\mid+C.$ 12. $\frac{1}{4}x^4+\frac{2}{3}x^{3/2}+ln\mid x\mid+C.$

13. $ln\mid x-1\mid+C.$ 14. $\tan x+ln\mid x+3\mid+C.$

15. $2\sqrt{x}-5\,ln\mid x+9\mid+ln\mid x-11\mid+C.$

16. $-3\cos x+8\sin x+C.$ 17. $11\sin x+5\tan x-4x+C.$

18. $-15\csc x+\frac{1}{7}\tan x-9\cos x+C.$

19. $5\tan x+C.$ 20. $-4\cot x+C.$

21. $\tan x-x+C.$ 22. $-\cot x-x+C.$

23. $x-\cos x+C.$ 24. $-\cot x+\csc x+C.$

25. $x+\cos x+C.$ 26. $\frac{5^x}{ln\,5}+C.$

27. $e^x+\frac{2^x}{ln\,2}+C.$ 28. $x+2e^x+C.$

29. $e^{x+7}+C.$ 30. $\frac{3^x}{ln\,3}+7\tan x+5e^x+C.$

31. $11\sin^{-1}x+C.$ 32. $\frac{1}{6}x^6-5\sec^{-1}x+C.$

33. $-\frac{5}{3x^3}+6\tan x+\sec^{-1}x+C.$

34. $x+6\tan^{-1}x+C.$ 35. $\frac{1}{3}x^3+4\tan^{-1}x+C.$

Exercise 5 (c)

1. $\frac{1}{2}(5x^2+7x+3)^2+C.$ 2. $\frac{1}{2}(5x^3+7x^2+3)^2+C.$

3. $\frac{1}{2}e^{x^2}+C.$ 4. $e^{\tan x}+C.$

5. $\frac{1}{9}ln\mid 9x+1\mid+C.$ 6. $ln\mid\cos x+\sin x\mid+C.$

7. $ln\,(x^2+1)+C.$ 8. $ln\,(1+e^x)+C.$

9. $-\frac{1}{x+5}+C.$ 10. $\tan(x-3)+C.$

11. $\frac{1}{12}(3x-7)^4+C.$ 12. $\frac{9^{11x}}{11\ ln\ 9}+C.$

13. $\frac{1}{7}(7x^2+5)^{1/2}+C.$ 14. $-\frac{3}{4}(\cos x)^{4/3}+C.$

15. $\frac{5}{3}\sec^3 x+C.$ 16. $\tan x+C.$

17. $\frac{1}{3}\tan^{-1}(x^3)+C.$ 18. $\frac{1}{2}\ ln\ (e^{2x}+1)+C.$

19. $\frac{1}{5}\sin(5\ ln\ x-2)+C.$ 20. $\tan\left(\frac{x}{2}\right)+C.$

21. $ln\ (\ |\ ln \sin x\ |\ )+C.$

22. $-\frac{1}{4}ln\ (5\cos^2 x+3\sin^2 x)+C.$

23. $\frac{1}{4}\ ln\ |\sec(4x+11)|+C.$

24. $x-2\ ln\ |\sec x+\tan x|+\tan x+C.$

25. $5\ ln\left|\tan\frac{x}{2}\right|+C.$ 26. $7\ ln\ |\tan x|+C.$

27. $x(\cos 2)+(\sin 2)\ ln\ |\sin x|+C.$

28. $x\cos\alpha-\sin\alpha\ ln\ |\sec x|+C.$

29. $2\ ln\left|\tan\frac{1}{2}\sqrt{x}\right|+C.$ 30. $ln\ |\sqrt{(1+x^2)}+x|+C.$

31. $ln\ |x|+C.$ 32. $ln\ |\sec(ln\ x)|+C.$

33. $\frac{1}{3}\sin^{-1}\left(\frac{3x}{4}\right)+C.$ 34. $\frac{1}{6}\tan^{-1}\left(\frac{2x}{3}\right)+C.$

35. $\frac{1}{5}\sec^{-1}\left(\frac{x}{5}\right)+C.$ 36. $\frac{1}{5}\sin^{-1}(5x+4)+C.$

Exercise 5 (*d*)

1. $\frac{1}{2}\tan^{-1}(x/2)+C.$ 2. $\frac{1}{6}\ ln\left|\frac{3+x}{3-x}\right|+C.$

3. $\frac{1}{2}\tan^{-1}(2x)+C.$ 4. $\frac{1}{10}\ ln\left|\frac{x-5}{x+5}\right|+C.$

5. $\frac{1}{2}\tan^{-1}(2x-1)+C.$ 6. $\frac{1}{3}\ ln\left|\frac{2x-1}{2x+2}\right|+C.$

7. $\frac{1}{3}\ ln\left|\frac{2x-1}{2x+2}\right|+C.$ 8. $ln\left|\frac{x-2}{x-1}\right|+C.$

9. $\frac{3}{4}\ ln\ |2x^2+x+1|+\frac{1}{2\sqrt{7}}\tan^{-1}\left(\frac{4x+1}{\sqrt{7}}\right)+C.$

10. $\frac{3}{2}\ ln\ |\ x^2-x-2\ |\ +\frac{1}{2}\ ln\ |\ (x-2)/(x+1)\ |\ +C.$

11. $\frac{5}{8}\ ln\ |\ 3x^2+2x+1\ |\ -\frac{11}{3\sqrt{2}}\tan^{-1}\left(\frac{3x+1}{\sqrt{2}}\right)+C.$

12. $\frac{1}{2}\ ln\ |\ x^2+2x+3\ |\ -(1/\sqrt{2})\tan^{-1}((x+1)/\sqrt{2})+C.$

Exercise 5 (*e*)

1. $\sin^{-1}\frac{x}{3}+C.$ 2. $ln\ |\ \sqrt{x^2+16}+x\ |\ +C.$

3. $\frac{1}{2}\sin^{-1}(2x)+C.$ 4. $\frac{1}{3}\ ln\ |\ \sqrt{9x^2+1}+3x\ |\ +C.$

5. $ln\ |\ \sqrt{x^2-16}+x\ |\ +C.$ 6. $\frac{1}{2}\ \ln\ |\ \sqrt{4x^2-9}+2x\ |\ +C.$

7. $\frac{1}{4}\sin^{-1}\left(\frac{4x}{3}\right)+C.$

8. $\frac{1}{5}ln\ |\ \sqrt{25x^2+16}+5x\ |\ +C.$

9. $\frac{1}{2}\sin^{-1}(x^2)+C.$ 10. $ln\ |\ \sqrt{e^{2x}+1}+e^x\ |\ +C.$

11. $\frac{1}{3}ln\ |\ \sqrt{x^6+1}+x^3\ |\ +C.$ 12. $ln\ |\ \sqrt{x^2+4x}+x+2\ |\ +C.$

13. $\frac{1}{\sqrt{2}}ln\left|\sqrt{x^2+\frac{3}{2}x+2}+\left(x+\frac{3}{4}\right)\right|+C.$

14. $ln\left|\left(x+\frac{1}{2}\right)+\sqrt{(x^2+x+1)}\right|+C.$

15. $\frac{1}{\sqrt{3}}ln\left|\left(x-\frac{1}{6}\right)+\sqrt{\left(x^2-\frac{1}{3}x+\frac{2}{3}\right)}\right|+C.$

16. $-\sqrt{(8+x-x^2)}+\frac{1}{2}\sin^{-1}\frac{2x-1}{\sqrt{33}}+C.$

17. $\sqrt{(x^2-x+1)}+\frac{3}{2}\ ln\ |\ \sqrt{(x^2-x+1)}+x-1\ |\ +C.$

18. $2\sqrt{(x^2+3x+1)}+2\ ln\left|\ \sqrt{(x^2+3x+1)}+x+\frac{3}{2}\right|+C.$

Exercise 5 (*f*)

1. $-x\cos x+\sin x+C.$ 2. $\frac{1}{2}x^2\ ln\ x-\frac{1}{4}x^2+C.$

3. $x\tan^{-1}x-\frac{1}{2}\ ln\ (1+x^2)+C.$

4. $x^2\sin x+2x\cos x-2\sin x+C.$

5. $\left(\frac{1}{2}x-\frac{1}{4}\right)e^{2x}+C.$ 6. $-(x^2+2x+2)e^{-x}+C.$

7. $x\sin^{-1}x+(1-x^2)^{1/2}+C.$

8. $\left(\frac{1}{8}-\frac{1}{4}x^2\right)\cos 2x+\frac{1}{4}x\sin 2x+C.$

9. $\frac{1}{2}(x^2-1)\ln(1+x)-\frac{1}{4}x^2+\frac{1}{2}x+C.$

10. $x\tan x-\ln|\sec x|+C.$ 11. $-x\cot\frac{1}{2}x+C.$

12. $-(x^2+3x+2)\cos x+(2x+3)\sin x+C.$

13. $\ln x\ln(\ln x)-\ln x+C.$

14. $-\sqrt{(1-x^2)}\sin^{-1}x+x+C.$

15. $x[(\ln x)^2-2\ln x+2]+C.$

16. $\frac{1}{4}x^2[2(\ln x)^2-2\ln x+1]+C.$

17. $\frac{1}{2}(x^2+1)\tan^{-1}x-\frac{1}{2}x+C.$

18. $x\cos^{-1}x-\sqrt{1-x^2}+C.$

19. $(x+1)\tan^{-1}\sqrt{x}-\sqrt{x}+C.$

20. $2[\sqrt{x}\sin^{-1}\sqrt{x}+\sqrt{(1-x)}]+C.$

21. $x(1-x^2)^{-\frac{1}{2}}\sin^{-1}x+\frac{1}{2}\ln(1-x^2)+C.$

22. $(1+x^2)\tan^{-1}x-\frac{1}{2}(\tan^{-1}x)^2+C.$

23. $\frac{1}{(\sqrt{1+x^2})}(x-\tan^{-1}x)+C.$ 24. $e^x(x-1)/(x+1)+C.$

25. $e^x\sin x+C.$ 26. $e^x\ln|\sec x|+C.$

27. $e^x\tan(x/2)+C.$ 28. $e^x\ln|\sin x|+C.$

29. $-e^x\cot(x/2)+C.$ 30. $e^x\ln|x|+C.$

31. $\frac{e^{5x}}{\sqrt{61}}\cos\left(6x-\tan^{-1}\frac{6}{5}\right)+C.$

32. $\frac{e^x}{\sqrt{26}}\sin(5x-\tan^{-1}5)+C.$

33. $\frac{1}{2}e^x\left[1-\frac{1}{\sqrt{5}}\cos(2x-\tan^{-1}2)\right]+C.$

34. $\frac{1}{2\sqrt{5}}e^x\sin(2x-\tan^{-1}2)+C.$

35. $\frac{x}{\sqrt{2}}\sin(\ln x-\pi/4)+C.$

Exercise 5 (g)

1. $\frac{1}{2}x\sqrt{(9-x^2)}+\frac{9}{2}\sin^{-1}(x/3)+C.$

2. $\frac{1}{2}x\sqrt{(x^2+7)}+\frac{7}{2}\ln|\sqrt{(x^2+7)}+x|+C.$

3. $\frac{1}{2}x\sqrt{(4x^2+9)}+\frac{9}{4}\ln|\sqrt{(4x^2+9)}+2x|+C.$

4. $\frac{1}{2}x\sqrt{(16-25x^2)}+\frac{8}{5}\ln|\sqrt{(16-25x^2)}+5x|+C.$

5. $\frac{1}{2}x\sqrt{(x^2-25)}-\frac{25}{2}\ln|\sqrt{(x^2-25)}+x|+C.$

6. $\frac{1}{2}x\sqrt{(16x^2-49)}-\frac{49}{8}\ln|\sqrt{(16x^2-49)}+4x|+C.$

7. $\frac{1}{2}x\sqrt{(4x^2-7)}-\frac{7}{4}\ln|\sqrt{(4x^2-7)}+2x|+C.$

8. $\frac{1}{4}x^2\sqrt{(x^4+1)}+\frac{1}{4}\ln(\sqrt{(x^4+1)}+x^2)+C.$

9. $\frac{1}{\sqrt{3}}\sin^{-1}\{(3x-1)/\sqrt{2}\}+C.$

10. $\frac{1}{6}(3x-2)\sqrt{(3x^2-4x+1)}-\frac{\sqrt{3}}{8}\ln\left|\sqrt{(3x^2-4x+1)}+\frac{3x-2}{\sqrt{3}}\right|+C.$

11. $\frac{1}{24}(8x^2-2x-11)\sqrt{(1+x-x^2)}+\frac{5}{16}\sin^{-1}\frac{2x-1}{\sqrt{5}}+C.$

12. $\frac{1}{6}(4x^2-18x+1)\sqrt{(2+3x-x^2)}-\frac{17}{2}\sin^{-1}\frac{2x-3}{\sqrt{(17)}}+C.$

Exercise 5 (h)

1. $\frac{1}{2}\ln|x+3|+\frac{3}{2}\ln|x+5|+C.$

2. $-4\ln|x-3|+3\ln|x-2|+C.$

3. $\frac{7}{16}\ln|2x-1|-\frac{37}{8}\ln|2x-3|+\frac{169}{16}\ln|2x-5|+C.$

4. $3\ln|x+1|+2\ln|x-2|+4\ln|x+3|+C.$

5. $3x+\frac{13}{4}\ln|x-3|-\frac{9}{4}\ln|x+1|+C.$

6. $x^2+3x+\frac{5}{2}\ln|2x+3|+\ln|x-1|+C.$

7. $\frac{1}{2}x^2-3x+4\ln|x+1|+3\ln|x-1|-7\ln|x+3|+C.$

8. $\frac{1}{2}x^2+3x+4\ln|x-1|+3\ln|x-2|+C.$

9. $\frac{1}{2}x+\frac{35}{36}\ln|2x-1|-\frac{10}{3}\ln|x+1|+\frac{1}{9}\ln|x-5|+C$

10. $\frac{2}{\sqrt{3}}\tan^{-1}(x/\sqrt{3})-\frac{1}{\sqrt{2}}\tan^{-1}(x/\sqrt{2})+C.$

11. $\frac{67}{6}\tan^{-1}(x/2)+\frac{7}{6}\tan^{-1}x-\frac{39}{2\sqrt{3}}\tan^{-1}(x/\sqrt{3})+C.$

12. $x+\frac{a^2}{a-b}\ln|x-a|+\frac{b^2}{b-a}\ln|x-b|+C.$

13. $\Sigma\frac{a^2}{(a-b)(a-c)}\ln|x-a|+C.$

14. $\Sigma\frac{pa^2+qa+r}{(a-b)(a-c)}\ln|x-a|+C.$

15. $x+\frac{(c-a)(c-d)}{c-d}\ln|x-c|+\frac{(d-a)(d-b)}{d-c}\ln|x-d|+C.$

16. $\frac{1}{2(a^2-b^2)}\{a\ln|\frac{x-a}{x+a}|-b\ln|\frac{x-b}{x+b}|\}+C.$

Exercise 5 (*i*)

1. $\frac{1}{4}\ln|\frac{x-1}{x+1}|-\frac{1}{2(x+1)}+C.$

2. $\frac{2}{5}\ln|x+3|+\frac{3}{5}\ln|x-2|-\frac{1}{x-2}+C.$

3. $\frac{1}{4}\ln|x+1|+\frac{3}{4}\ln|x-1|-\frac{3}{2}\frac{1}{x-1}+C.$

4. $\ln|(x-1)(x+3)|-\frac{1}{x-1}+C.$

5. $\frac{3}{5}\ln|x-2|+\frac{2}{5}\ln|x+3|-\frac{3}{x-2}+C.$

6. $\frac{5}{x+1}-\frac{1}{(x+1)^2}+\frac{7}{2}\ln|x+1|+\frac{1}{2}\ln|x-1|+C.$

7. $\frac{4}{27}\ln|x-1|-\frac{4}{27}\ln|x+2|-\frac{1}{6(x-1)^2}-\frac{5}{9(x-1)}+C.$

8. $\frac{3}{16}\ln|x-1|-\frac{3}{16}\ln|x+1|-\frac{5}{8}\frac{1}{(x+1)}+\frac{3}{8}\frac{1}{(x+1)^2}$

$-\frac{1}{6}\frac{1}{(x+1)^3}+C.$

9. $\frac{1}{3}\frac{1}{(x-2)}-\frac{1}{(x-2)^2}+\frac{1}{9}\ln|\frac{x+2}{x-1}|+C.$

10. $\frac{2}{x-1}+\frac{1}{2}\frac{1}{(x-1)^2}+2\ln\left|\frac{x-2}{x-1}\right|+C.$

11. $-\frac{1}{9}\frac{1}{(x-1)}-\frac{4}{3}\frac{1}{(x-1)^2}+\frac{1}{27}\ln\left|\frac{x+2}{x-1}\right|+C.$

12. $-\frac{1}{2(x-1)}+\ln|x|-\frac{1}{4}\ln|x+1|-\frac{3}{4}\ln|x-1|+C.$

13. $-\frac{1}{2(x+1)}+\frac{3}{4}\ln|x+1|-\ln|x|+\frac{1}{4}\ln|x-1|+C.$

14. $-\frac{1}{27}\frac{1}{(x+1)}-\frac{20}{27}\frac{1}{(x-2)}-\frac{4}{9}\frac{1}{(x-2)^2}+\frac{2}{27}\ln\left|\frac{x-2}{x+1}\right|+C.$

15. $\frac{a}{x+a}+\frac{1}{2}\ln|x^2-a^2|+C.$

Exercise 5 (*j*)

1. $\frac{1}{3}\ln|x-1|-\frac{1}{3}\ln|x^2+x+1|$

$$+\frac{4}{3\sqrt{3}}\tan^{-1}\{(2x+1)/\sqrt{3}\}+C.$$

2. $\frac{1}{8}\ln|x-2|-\frac{1}{16}\ln(x^2+4)+\frac{3}{8}\tan^{-1}\frac{x}{4}+C.$

3. $-\ln|x-1|+\frac{1}{2}\ln(x^2+2)+C.$

4. $\ln(x^2+1)-2\ln|x+3|+3\tan^{-1}x+C.$

5. $\frac{2}{3}\ln|x+1|+\frac{1}{6}\ln|x^2-x+1|$

$$+\frac{2}{3\sqrt{3}}\tan^{-1}\{(2x-1)/\sqrt{3}\}+C.$$

6. $-2\ln|5-3x|+\ln(4x^2+3)+\frac{1}{2\sqrt{3}}\tan^{-1}(2x/\sqrt{3})+C.$

7. $\frac{1}{2}x^2+\frac{1}{4}\ln|x^2-1|-\frac{1}{4}\ln(x^2+1)+C.$

8. $\frac{7}{60}\ln\left|\frac{x-3}{x+3}\right|+\frac{3}{10}\tan^{-1}x+C.$

9. $\frac{1}{9}\ln|x-1|+\frac{2}{9}\ln|x+2|-\frac{1}{6}\ln|x^2+x+1|$

$$-\frac{1}{3\sqrt{3}}\tan^{-1}\{(2x+1)/\sqrt{3}\}+C.$$

10. $\frac{3}{5}\ln|x-1|-\frac{1}{4}\ln|x-2|-\frac{7}{40}\ln(x^2+4)$

$$-\frac{1}{10}\tan^{-1}x+C.$$

11. $-\frac{1}{2}\ln|x-1|+\frac{1}{10}\ln|x+1|-\frac{1}{5}\ln(x^2+4)+$

$$\frac{1}{5}\tan^{-1}(x/2)+C.$$

12. $\frac{1}{2}\ln|x+1|-\frac{1}{4}\ln(x^2+1)-\frac{1}{2(x+1)}+C.$

13. $\frac{1}{2}\ln|x-1|-\frac{1}{6}\ln|x+1|-\frac{1}{6}\ln|x^2-x+1|$

$$+\frac{1}{\sqrt{3}}\tan^{-1}\frac{2x-1}{\sqrt{3}}+C.$$

Exercise 5 (k)

1. $(2/\sqrt{5})\tan^{-1}\left\{(1/\sqrt{5})\tan\left(\frac{1}{2}x\right)\right\}+C.$

2. $(2/\sqrt{3})\tan^{-1}\left\{(1/\sqrt{3})\tan\left(\frac{1}{2}x\right)\right\}+C.$

3. $\frac{1}{2}\tan^{-1}\left\{\frac{1}{2}\tan\left(\frac{1}{2}x\right)\right\}+C.$

4. $(1/\sqrt{3})\ln\left|\dfrac{\sqrt{3}+\tan\left(\frac{1}{2}x\right)}{\sqrt{3}-\tan\left(\frac{1}{2}x\right)}\right|+C.$

5. $(2/\sqrt{7})\tan^{-1}\left\{(1/\sqrt{7})\tan\left(\frac{1}{2}x\right)\right\}+C.$

6. $\frac{2}{5}\tan^{-1}\left\{\frac{1}{5}\tan\left(\frac{1}{2}x\right)\right\}+C.$

7. $\frac{1}{\sqrt{119}}\ln\left|\dfrac{\sqrt{17}+\sqrt{7}\tan\left(\frac{1}{2}x\right)}{\sqrt{17}-\sqrt{7}\tan\left(\frac{1}{2}x\right)}\right|+C.$

8. $\frac{1}{\sqrt{3}}\ln\left|\dfrac{\tan(x/2)-2-\sqrt{3}}{\tan(x/2)-2+\sqrt{3}}\right|+C.$

Exercise 5 (l)

1. $\sin x-\frac{2}{3}\sin^3 x+\frac{1}{5}\sin^5 x+C.$

2. $-\cos x+\cos^3 x-\frac{3}{5}\cos^5 x+\frac{1}{7}\cos^7 x+C.$

3. $-\frac{1}{4}\sin^3 x\cos x-\frac{3}{8}\sin x\cos x+\frac{3}{8}x+C.$

4. $\frac{1}{4}\cos^3 x \sin x+\frac{3}{8}\sin x\cos x+\frac{3}{8}x+C.$

5. $\frac{1}{3}\sin^3 x-\frac{1}{5}\sin^5 x+C.$

6, $\frac{1}{6}\sin^3\cos^3 x+\frac{1}{8}\sin^3 x\cos x+\frac{1}{16}(x-\sin x\cos x)+C.$

7. $-\frac{1}{6}\cos^6 x+\frac{1}{8}\cos^8 x+C.$

8. $-\frac{1}{8}\cos^3 x\sin^5 x-\frac{5}{48}\cos^3 x\sin^3 x-\frac{5}{64}\cos^3 x\sin x+\frac{5}{128}(x+\sin x\cos x)+C.$

9. $-\frac{1}{8}\sin^3 x\cos^5 x-\frac{1}{16}\sin x\cos^5 x+\frac{1}{64}\sin x\cos^3 x+\frac{3}{128}(x+\sin x\cos x)+C.$

10. $-\frac{1}{7}\cos^7 x+\frac{2}{9}\cos^9 x-\frac{1}{11}\cos^{11} x+C$

Test Your Understanding V

1. (*c*). 2. (*d*). 3. (*d*). 4. (*c*). 5. (*a*).
6. (*b*). 7. (*a*). 8. (*a*). 9. (*b*). 10. (*b*).

Review Exercise V

1. $x-\ln|x|+\frac{1}{2}\ln(x^2+1)-\tan^{-1}x+C.$

2. $\ln(x+1)+\frac{1}{6}\ln x^{2/3}-x^{1/3}+1^{1/3})+\frac{1}{3}\ln|x^{1/3}+1|-\frac{1}{\sqrt{3}}\tan^{-1}\frac{2x^{1/3}-1}{\sqrt{3}}+C.$

3. $-\frac{1}{2}\cdot\frac{1}{x^2}\ln x-\frac{1}{4}\cdot\frac{1}{x^2}+C.$

4. $\frac{1}{2}x^2+x+\frac{1}{2}\ln|x-1|-\frac{1}{4}\ln(x^2+1)+\frac{1}{2}\tan^{-1}x+C.$

5. $\frac{1}{9}\{(2x+3^{3/2})-(2x)^{3/2}\}+C.$

6. $x+\ln\left|\frac{x-2}{x+3}\right|+C.$

7. $\sqrt{(a^2-x^2)}+a\sin^{-1}(x/a)+C.$

8. $\frac{1}{7b}(a+be^x)^7+C.$
9. $-x+3 \ln |x+2|+2 \ln |x-1|+C.$
10. $-e^x \cot x+C.$
11. $\sin^{-1} x^2+C.$
12. $\ln |\cos x+\sin x|+C.$
13. $\ln |1+\sin x|+C.$
14. $3 [x \tan^{-1} x-\frac{1}{2} \ln (1+x^2)]+C.$
15. $-\frac{2}{1+\tan (x/2)}+C.$
16. $\frac{2}{3(a-b)}[(x+a)^{3/2}-(x+b)^{3/2}]+C.$
17. $\tan (1+\ln x)+C.$
18. $\ln |\sec x(\sec x+\tan x)|+C.$
19. $x[(\ln x)^2-2(\ln x)+2]+C.$
20. $\sin^{-1} \frac{2x^2+1}{\sqrt{5}}+C.$
21. $e^x \ln |\sec x|+C.$
22. $4 \sin x-\frac{8}{3} \sin^3 x+C.$
23. $\ln \frac{x^2+1}{x^2+2}+C.$
24. $\frac{1}{b^2-a^2} \ln (a^2 \cos^2 x+b^2 \sin^2 x)+C.$
25. $2(1+\cos \theta)^{1/2}+C.$
26. $-\frac{1}{24} \cos 12x-\frac{1}{4} \cos 2x+C.$
27. $\frac{1}{2}[\ln (\sec x+\tan x)]^2+C.$
28. $e^x \cot x+C.$
29. $\frac{1}{2}\left(\ln |2x+1|+\frac{1}{2x+1}\right)+C.$
30. $\frac{1}{2\sqrt{5}} \tan^{-1}\left(\frac{2}{\sqrt{5}} \tan x\right)+C.$

Exercise 6 (*a*)

1. $\frac{1}{2}(b^2-a^2).$ 2. $\frac{3}{2}(b^2-a^2)+5(b-a).$

3. $\frac{1}{3}(b^3-a^3)$. 4. $\frac{1}{4}(b^4-a^4)$.

5. $\sin b-\sin a$. 6. $\frac{1}{2}(\cos 2a-\cos 2b)$.

7. $\frac{1}{2}$. 8. $e^{-a}-e^{-b}$.

9. $\pi/4$. 10. $\frac{1}{2}(e-1)$.

Exercise 6 (*b*)

1. 64. 2. 40·5.
3. $ln\ 2$. 4. $e-1$.
5. $1/\sqrt{3}$. 6. $\frac{83}{6}$.
7. $\pi/6$. 8. $\sqrt{2}-1$.
9. $\frac{1}{2}\ ln\ 2$. 10. $\pi/8$.
11. $\pi/8$. 12. $(\pi-4)/2$.
13. -2. 14. $\frac{1}{4}$.
15. $\frac{1}{4}(e^2-1)$. 16. $\frac{\pi}{3\sqrt{3}}$.
17. $\frac{2}{3}$. 18. $\frac{3}{16}\pi$.
19. $\pi/32$. 20. $\pi/48$.

Exercise 6 (*c*)

1. $\frac{7}{3}$. 2. $\frac{1}{3}\ ln\ 2$.
3. $\frac{1}{3}(5\sqrt{5}-8)$. 4. $\frac{2}{3}$
5. $6-4\ ln\ 2$. 6. $\tan^{-1} e-\pi/4$.
7. $\sin(ln\ 3)$. 8. $\frac{1}{4}\ ln\ \frac{82}{81}$.
9. $\pi/6$. 10. $\frac{\pi^{3/2}}{12}$
11. $\frac{1}{3}\ ln\ \frac{16}{9}$. 12. $\frac{\pi}{2\sqrt{2}}$.
13. $\frac{\pi}{4}\left(\frac{\pi}{4}-1\right)+\frac{1}{2}\ ln\ 2$. 14. $\frac{1}{10}\ ln\ 2$.

15. $\frac{\pi}{2}-1$.

16. $\pi/4-\frac{1}{2}\ ln\ 2$.

17. $\frac{2}{\sqrt{7}}\tan^{-1}\left(\frac{1}{\sqrt{7}}\right)$.

18. $\frac{1}{3}\ ln\ 2$.

19. $\pi/(1-a^2)$.

20. $ln\ (4/3)$.

21. $\pi/4$.

Exercise 6 (*d*)

1. $ln\ 2$.

2. 3/8.

3. $\pi/4+\frac{1}{2}\ ln\ 2$.

4. $\pi/4$.

5. $ln\ 3$.

6. $\pi/2$.

7. 1/3.

8. $\frac{1}{3}\ ln\ 2$.

9. $\frac{1}{10}(5-\sqrt{5})$.

10. $1/\sqrt{2}$.

11. 1/14.

12. 1/105.

13. $\pi/4$.

14. $\frac{1}{2}\pi+1$.

15. 1/24.

Exercise 6 (*f*)

1. 38·5.

2. e^3-1.

3. 4.

4. $2\ ln\ 2-1$.

5. πa^2.

6. $\frac{8}{3}a^2$.

7. 4.

8. 253/12.

9. $\frac{4}{3}$.

10. $\left(\frac{\pi}{4}-\frac{2}{3}\right)a^2$.

11. $\frac{16}{3}a^2$.

12. $\frac{1}{3}+\frac{5}{2}\left(\frac{\pi}{2}-\sin^{-1}(1/\sqrt{5})\right)$

13. $\frac{8}{3}$.

14. $\frac{9}{16}$.

15. 3π.

16. $\frac{16}{3}$.

17. $\frac{4}{3}$.

18. $\frac{16}{3}\sqrt{2}$.

19. $\frac{128}{9}$

20. $\frac{16}{3}$.

21. $\frac{56}{9}$.

22. $\frac{16}{3}$.

23. 6. 24. $\pi/4$.

25. $\frac{8}{3}$. 26. $\frac{27}{4}$.

27. $\frac{a^2}{3}$.

Test Your Understanding VI

1. (*d*). 2. (*c*). 3. (*a*). 4. (*c*). 5. (*d*).
6. (*b*). 7. (*c*). 8. (*a*). 9. (*b*). 10. (*c*).

Review Exercise VI

1. $\pi/2$. 2. 0.
3. 4. 4. $\pi/4$.
5. $\frac{1}{36}(\pi^2+16)$. 6. $a\pi$.
7. $\pi/8$. 8. 1.
9. 1. 10. $\frac{\pi}{4}-\frac{1}{2}\ln 2$.
11. $\frac{1}{3}(e^6-e^3)$. 12. (*a*) $\sqrt{2}\tan^{-1}\sqrt{2}-\frac{\pi}{4}$,
(*b*) $\frac{1}{105}$. 20. $\frac{1}{4}\pi ab$.
21. $|a|$. 22. 4π.
23. $\frac{11}{8}$ 24. $2\left(\pi-\frac{2}{3}\right)a^2$.
25. $\frac{1}{3}a^2$. 26. $\frac{3}{2}\pi$.
27. $\frac{7}{120}$.

Exercise 7 (*a*)

1. O=1, D=1. 2. O=1, D=1.
3. O=3, D=2. 4. O=4, D=2.
5. O=1, D=1. 10. $xy'-4y=0$.
11. 2. 12. $y'-1+(y-x)\tan x=0$.
13. $2xy'-y=0$. 14. $xy'+y=0$.
15. $(x+yy')(xy'-y)=2y'$.

Exercise 7 (*b*)

1. $y^2=2\sin x+C$. 2. $y^2(C-2x)=1$.
3. $x^2+3=Ce^{2y}$. 4. $v+2=C(u+1)$.
5. $(\cos x+C)e^y=1$. 6. $r^2=\sin 2\theta+C$.

7. $3y^4=4(t+1)^3+C$.   8. $\frac{1}{2y}+ln \mid \tan x \mid =C$.

9. $r^2=C\,e^{\theta+\sin\theta}$.   10. $y=C(x+4)$.

11. $e^x+e^{-y}=C$.   12. $x-e^{-x}=y-ln \mid 1+y \mid +C$.

13. $x+ln \mid x/(1-y) \mid =C$.   14. $y=4\tan(x/2)-2x+C$.

15. $(2-x^2)\cos x+2x\sin x+\frac{1}{2}(ln\, y)^2=C$.

16. $\cos x=C\sin y$.   17. $e^x\, ln\, x=\sin y+C$.

18. $\frac{1}{3}(x+a)^{1/2}(a-2x)+ln \mid y \mid =C$.

19. $(y-1)^2(1-x^2)=Cy^2$.   20. $xy=1$.

21. $27\,y=x(y+3)^3$.   22. $y=\sec x$.

23. $Y(t)$ is 8 times in 9 hrs,

25. 5·4.

27. $y=Ax^k$.   28. $x^3-3y-240=0$.

29. $(y+1)=\sqrt{2\,\frac{x-1}{x+1}}$.

30. $x^2+y^2=1$.

Exercise 7 (c)

1. $x^2+2xy=C$.

2. $\frac{1}{2}ln(x^2+y^2)+\arctan(y/x)=C$.

3. $\frac{1}{2}ln(x^2+y^2)-\arctan(y/x)=C$.

4. $x^2-y^2=Cx$.   5. $(y-x)^2=Cxy^2$.

6. $e^{x^2/2y^2}=Cy$.   7. $y/x=ln\, y+C$.

8. $x^3=3y^3\, ln(Cy)$.   9. $ln\, x=\cos(y/x)+C$.

10. $x\sin(y/x)=C$.   11. $x^2+2xy-y^2=2$.

Exercise 7 (d)

1. $y=e^{-x}+C$.   2. $y=Ce^x-x$.

3. $y=Ce^{4x}-x^2$.   4. $y=Ce^{-3x}+\frac{2}{5}e^{2x}+4$.

5. $y=Ce^{3x}+xe^{3x}$.   6. $y=(x+C)e^x$.

7. $y=Ce^{3x}-2\cos 2x-3\sin 2x$.

8. $y=Ce^{-2x}+2\cos x+\sin x$.

9. $y=(\sin^2 x+C)e^{4x}$.

10. $y=e^{-2x}(\cos x+\sin x)+Ce^{-3x}$.

Test Your Understanding VII

1. (*a*). 2. (*d*). 3. (*c*). 4. (*d*). 5. (*b*).
6. (*b*). 7. (*b*). 8. (*b*). 9. (*a*). 10. (*d*).

Review Exercise VII

1. $|1+y| = Ce^{x+\frac{1}{2}x^2}$.

2. $\frac{1}{2}(x^2+y^2)+x+y+\frac{1}{2}\ ln\ |(x-y)(y-1)| = C$.

3. $(1-e^x)^3 = C \tan y$.

4. $\tan^{-1} x + \tan^{-1} y + \frac{1}{2}\ ln\ [(1+x^2)(1+y^2)] = C$.

5. $\frac{2}{3}(x+a)^{1/2}(x-2a) + ln\ |y| = C$.

6. $\frac{1}{3}(2x+3y)+\frac{9}{64}\ ln\ |16x+24y+33| = x+C$.

7. $(y-x) \csc x = C$. 8. $(x+y)^3 = c(y-x)$.

9. $xy\sqrt{(y^2-x^2)} = C$. 10. $x^2+y^2 = c(x+y)$.

11. $y = Ce^{-x}+\frac{1}{2}x^2$. 12. $y = (c+\sin x \cos x)e^{2x}$.

13. $y = (x^2+C)e^{-5x}$. 14. $y = \sin 3x - \cos 3x + (x+C)e^{3x}$.

15. $y(1+ax)(1-a^2) = x(1-ay)(1+a^2)$.

16. 2000 units.

17. $\frac{1}{2}\ ln\ (x^2+1) + \arctan y = \pi/4$.

18. $y = x^2$.

Exercise 8 (*a*)

1. No. They represent equal and opposite vectors.
2. Yes. 3. Yes.
4. No, except when the trapezium is a parallelogram.
5. $\mathbf{a}+\mathbf{b}+\mathbf{c}$. 11. $\mathbf{a}+\mathbf{b}$, $\mathbf{b}-\mathbf{a}$, $2\mathbf{b}-2\mathbf{a}$.

Exercise 8 (*c*)

1. (*i*) 3. (*ii*) $-6\sqrt{2}$, (*iii*) $10\sqrt{3}$.
3. $2\sqrt{2}$. 4. (*i*) -1, (*ii*) 4, (*iii*) 15.
5. (*i*) 7, (*ii*) 9. 6. (*i*) 90°, (*ii*) 180°.
13. In a triangle ABC, $c^2 = a^2+b^2-2ab \cos C$.

Exercise 8 (*d*)

13. 9 units. 14. -15 units.

Exercise 8 (*e*)

1. (*i*) $7\mathbf{i}-10\mathbf{j}+\mathbf{k}$, (*ii*) $-7\mathbf{i}+10\mathbf{j}-\mathbf{k}$.
2. (*i*) $-\mathbf{i}-3\mathbf{j}-7\mathbf{k}$, (*ii*) $-\mathbf{i}-10\mathbf{j}+8\mathbf{k}$, (*iii*) $4\mathbf{i}-6\mathbf{j}-7\mathbf{k}$, (*iv*) $9\mathbf{i}-10\mathbf{j}+3\mathbf{k}$.
3. $\frac{1}{\sqrt{5}}\,\mathbf{j}+\frac{2}{\sqrt{5}}\,\mathbf{k}$.
4. 6.
5. 4.
7. $\pm\frac{1}{\sqrt{414}}\,(17\mathbf{i}+2\mathbf{j}-11\mathbf{k})$.
8. $\pm\frac{1}{\sqrt{3}}\,(-\mathbf{i}-\mathbf{j}+\mathbf{k})$.
11. $13\mathbf{i}+11\mathbf{j}-\mathbf{k}$, $5\mathbf{i}+5\mathbf{j}-10\mathbf{k}$. They are not equal.
13. $\pm\frac{1}{\sqrt{3}}\,(\mathbf{i}-\mathbf{j}+\mathbf{k})$.

Exercise 8 (*f*)

1. $5\sqrt{3}$ sq. units.
2. $5\sqrt{3}$ sq. units.
3. $3\sqrt{2}$ sq. units.
10. $\mathbf{i}+2\mathbf{j}+4\mathbf{k}$, magnitude $\sqrt{21}$.
11. $6\,(\mathbf{i}+\mathbf{j})$, magnitude $6\sqrt{2}$.
12. $-2\mathbf{i}+6\mathbf{j}-10\mathbf{k}$; magnitude $\sqrt{140}$.

Exercise 8 (*g*)

1. (*a*) 12, (*b*) -12.
2. 0.
4. (*a*) 5 cu. units, (*b*) 12 cu. units.
5. 100 cu. units.

Exercise 8 (*h*)

1. (*a*) $2\mathbf{i}+9\mathbf{j}-2\mathbf{k}$, (*b*) $-\mathbf{i}+9\mathbf{j}+6\mathbf{k}$.
2. (*a*) $26\mathbf{i}-\mathbf{j}+17\mathbf{k}$, (*b*) $-13\mathbf{i}-9\mathbf{j}+2\mathbf{k}$, (*c*) $-13\mathbf{i}+10\mathbf{j}-19\mathbf{k}$, (*d*) $13\mathbf{i}-10\mathbf{j}+19\mathbf{k}$, (*e*) $-26\mathbf{i}+\mathbf{j}-17\mathbf{k}$, (*f*) $13\mathbf{i}+9\mathbf{j}-2\mathbf{k}$,
12. $\pi/2$, $\pi/3$.

Test Your Understanding VIII

1. (*d*). 2. (*d*). 3. (*c*). 4. (*a*). 5. (*b*).
6. (*a*). 7. (*c*). 8. (*d*). 9. (*d*). 10. (*b*).

Review Exercise VIII

1. $\cos^{-1} 2/5\sqrt{202}$.
2. 8.
3. $\pm(5\mathbf{i}-\mathbf{j}-5\mathbf{k})$.
4. -2.
5. 5/2.
8. $\pm\frac{1}{\sqrt{59}}\,(3\mathbf{i}-\mathbf{j}+7\mathbf{k})$.
9. 16.
10. 6.

11. $\pm\dfrac{1}{\sqrt{3}}(\mathbf{i}-\mathbf{j}+\mathbf{k})$.

12. 13 units, 17 units, 41 units ; 455 sq. units, 435 sq. units, 320 sq. units ; 3396 cu. units.

13. -4.

15. $-15\mathbf{i}+10\mathbf{j}-9\mathbf{k},\ 3\mathbf{i}+5\mathbf{j}-\mathbf{k}$.

16. (*i*) -15, (*ii*) $2/3$.

17. $-11\mathbf{i}+122\mathbf{j}-85\mathbf{k},\ -46\mathbf{i}+66\mathbf{j}+34\mathbf{k}$.

Exercise 9 (*a*)

1. (*a*) $\left(\dfrac{2}{3}, \dfrac{1}{3}, \dfrac{2}{3}\right)$, (*b*) $\left(-\dfrac{2}{3}, \dfrac{1}{3}, \dfrac{2}{3}\right)$,

   (*c*) $\left(\dfrac{2}{3}, -\dfrac{1}{3}, \dfrac{2}{3}\right)$, (*d*) $\left(\dfrac{2}{3}, \dfrac{-2}{3}, \dfrac{1}{3}\right)$.

2. (*a*) $\left(-\dfrac{3}{5\sqrt{2}}, \dfrac{4}{5\sqrt{2}}, \dfrac{1}{\sqrt{2}}\right)$, (*b*) $\left(\dfrac{1}{\sqrt{2}}, \dfrac{-4}{5\sqrt{2}}, \dfrac{3}{5\sqrt{2}}\right)$,

   (*c*) $\left(-\dfrac{1}{\sqrt{2}}, \dfrac{3}{5\sqrt{2}}, \dfrac{4}{5\sqrt{2}}\right)$, (*d*) $\left(\dfrac{3}{5\sqrt{2}}, \dfrac{-1}{\sqrt{2}}, \dfrac{4}{5\sqrt{2}}\right)$.

3. (*a*) (1, 0, 0), (*b*) (0, 1, 0),
   (*c*) (0, 0, 1), (*d*) (−1, 0, 0),
   (*e*) (0, −1, 0), (*f*) (0, 0, −1).

4. (*a*) $\pi/4$, (*b*) $\cos^{-1}\left(-\dfrac{1}{3}\right)$.

Exercise 9 (*b*)

1. (*a*) 13 units, 17 units, $\sqrt{38}$ units.
   (*b*) $\sqrt{120}$ units, $5\sqrt{2}$ units, $\sqrt{69}$ units.

2. (*a*) $\sqrt{254}$ units, (*b*) $\sqrt{69}$ units
   (*c*) $\sqrt{59}$ units.

Exercise 9 (*c*)

1. (−2, 1, −3).

2. $\left(\dfrac{9}{4}, -\dfrac{5}{4}, -\dfrac{1}{2}\right)$.

3. (−37, 67, 0).

4. (8, 21, −14).

5. 4 : −3.

6. 3 : −2.

7. (4, −5, 5).

8. (1, 2, 2).

Exercise 9 (*d*)

1. $\mathbf{r}=(2\mathbf{i}+3\mathbf{j}-\mathbf{k})+t(\mathbf{i}-\mathbf{j}+2\mathbf{k})$.

2. $\mathbf{r}=(\mathbf{i}-\mathbf{j}+2\mathbf{k})+t(2\mathbf{i}+\mathbf{j}+2\mathbf{k})$.

3. $\dfrac{x+1}{1}=\dfrac{y-1}{2}=\dfrac{z-1}{-2}$.

4. $\mathbf{r}=(\mathbf{i}-\mathbf{j}-2\mathbf{k})+t(2\mathbf{i}-3\mathbf{j}+4\mathbf{k})$.

5. $\dfrac{x-2}{0}=\dfrac{y-3}{1}=\dfrac{z}{-2}$.

6. $\mathbf{r}=(\mathbf{i}-\mathbf{j}+\mathbf{k})+t(\mathbf{i}+2\mathbf{j}-4\mathbf{k})$.

7. $\mathbf{r}=(2\mathbf{i}+\mathbf{j}+3\mathbf{k})+t(-3\mathbf{i}+\mathbf{j}-8\mathbf{k})$.

8. $\dfrac{x-2}{-5}=\dfrac{y+1}{2}=\dfrac{z+3}{7}$.

9. $\dfrac{x+1}{4}=\dfrac{y+1}{2}=\dfrac{z-2}{1\frac{2}{3}}$

10. $\dfrac{x-1}{1}=\dfrac{y+1}{-2}=\dfrac{z-1}{3}$

Exercise 9 (*e*)

1. $\cos^{-1}\dfrac{8}{5\sqrt{29}}$.
2. $\cos^{-1}\dfrac{1}{\sqrt{39}}$.
4. $\cos^{-1}\dfrac{22}{\sqrt{29}\sqrt{41}}$.
6. $\dfrac{\sqrt{70}}{14}$; Yes.
8. (1, 3, 2).
9. (2, 6, 3).
10. $\dfrac{x-3}{1}=\dfrac{y+1}{-6}=\dfrac{z-11}{4}$.
11. $2\sqrt{29}$, $\dfrac{1}{2}(x-1)=\dfrac{1}{3}(y-2)=\dfrac{1}{4}(z-3)$.
12. (3, 5, 7), (11, 11, 31).
13. $\sqrt{15}$; $x-4=\dfrac{(y-2)}{3}=-\dfrac{(z+3)}{5}$.
14. $4\sqrt{3}$: $x=y=z$.
15. 14 ; $\dfrac{x-5}{5}=\dfrac{y-7}{3}=\dfrac{z-3}{6}$; (8, −9, 10), (15, 29, 5).

Exercise 9 (*f*)

1. $\mathbf{r}.\mathbf{j}=2$.
2. $\mathbf{r}.(\mathbf{i}+2\mathbf{j}-2\mathbf{k})=15$.
3. 2.
4. $3x+4y-5z+4=0$.
5. $x-2y+2z=9$.
6. $\mathbf{r}.(2\mathbf{i}+\mathbf{j}+\mathbf{k})=6$.
7. $x+y+z=2$.
8. $\mathbf{r}.(\mathbf{i}+\mathbf{k})=1$, $x+z=1$.
9. $5x+2y-3z=17$
10. $3x-4z+1=0$.
11. $y+4z=7$
12. $4x-3y+2z=3$.

Exercise 9 (*g*)

1. $\dfrac{\pi}{2}$.
2. $\dfrac{\pi}{3}$.
3. $\dfrac{\pi}{2}$.
4. 2.

5. $\frac{1}{6}$. 6. $28x-17y+9z=0$.

7. $x-10y-5z=0$. 8. $5x-y+z=14$.

9. $x+7y+13z+96=0$.

10. $2x+y-2z+3=0$, $x-2y-2z-3=0$.

11. $y-3z+6=0$. 15. $4x-y-2z-6=0$.

16. $6x+y-16=0$.

17. $(\mathbf{r}.\mathbf{n}_1-1)(\mathbf{n}_2 . \mathbf{n}_3)=(\mathbf{r} . \mathbf{n}_2-1)(\mathbf{n}_1 . \mathbf{n}_3)$.

18. $[\mathbf{r}\ \mathbf{n}_1\ \mathbf{n}_2]=[\mathbf{a}\ \mathbf{n}_1\ \mathbf{n}_2]$. 19. $\mathbf{r} . (\mathbf{b}-\mathbf{a})\times\mathbf{n}=[\mathbf{a}\ \mathbf{b}\ \mathbf{n}]$.

Exercise 9 (*h*)

1. (*i*) $x^2+y^2+z^2-2x-3=0$.
   (*ii*) $x^2+y^2+z^2-2y+2z-7=0$.
   (*iii*) $x^2+y^2+z^2-4x+2y-6z-11=0$.

2. (*i*) $|\mathbf{r}-3\mathbf{j}|=2$. (*ii*) $|\mathbf{r}-(\mathbf{i}-\mathbf{j})|=4$.
   (*iii*) $|\mathbf{r}-(\mathbf{i}+\mathbf{j}-\mathbf{k})|=8$.

3. $x^2+y^2+z^2-6x-4y+2z-22=0$.

4. $|\mathbf{r}-(\mathbf{i}+2\mathbf{j}+3\mathbf{k})|=4$.

5. (*i*) Centre (0, 0, 0), radius$=5$ units,
   (*ii*) Centre (1, -2, 0), radius$=\sqrt{5}$ units,
   (*iii*) Centre (-3, 4, -5), radius$=8$ units,
   (*iv*) Centre (-2, -3, 4), radius$=6$ units.

6. Centre $3\mathbf{k}$, radius$=3\sqrt{2}$ units.

7. Centre (4, -3, 5), radius$=10$ units.

8. $x^2+y^2+z^2=14$.

9. $(\mathbf{r}-\mathbf{i}-\mathbf{j}).(\mathbf{r}-\mathbf{j}-\mathbf{k})=0$. Centre $\mathbf{i}+\frac{1}{2}\mathbf{j}+\frac{1}{2}\mathbf{k}$, radius$=\sqrt{2}/2$.

10. $x^2+y^2+z^2-2x-4y+4z+3=0$, centre (1, 2, -2), radius$=\sqrt{6}$ units.

11. (4, -2, 1).

12. Centre $\left(-2, \frac{3}{5}, \frac{-4}{5}\right)$, radius$=2$ units.

13. $3x^2+3y^2+3z^2-4x-4y-4z+1+0$.

14. Centre $\left(-\frac{1}{2}, -1, -\frac{3}{2}\right)$, radius$=\sqrt{126}$.

15. $3x^2+3y^2+3z^2+10x-2y-2z-42=0$.

16. $x^2+y^2+z^2-4x-6y-2z+2=0$.

17. $4x^2+4y^2+4z^2-8x-16y-24z+31=0$.

18. Centre $\left(\frac{1}{2}, -\frac{3}{4}, 1\right)$, radius$=\frac{1}{4}\sqrt{37}$.

Test Your Uunderstanding IX

1. (*c*). **2.** (*d*). **3.** (*b*). **4.** (*b*). **5.** (*a*) **6.** (*b*).
7. (*c*). **8.** (*c*). **9.** (*c*). **10.** (*a*).

Review Exercise IX

1. $\left(\frac{6}{7}, \frac{2}{7}, \frac{3}{7}\right)$. **2.** $\frac{29}{7}$.

3. $(-3, -1, -1)$. **4.** $(-18, 30, -8)$.

5. $\frac{x-2}{6}=\frac{y-4}{3}=\frac{z+1}{2}$; $(-4, 1, -3)$.

6. $\left(2, \frac{7}{2}, 50\right), \left(0 \frac{1}{2} 1,\right)$.

7. $\text{cso}^{-1}\left(\frac{17}{\sqrt{660}}\right), \cos^{-1}\left(\frac{21}{5\sqrt{77}}\right), \cos^{-1}\left(\frac{7}{\sqrt{420}}\right)$.

10. 12. **11.** $\left(\frac{11}{3}, -2, \frac{7}{3}\right)$.

12. 1. **13.** $\frac{15}{16}$.

14. 1. **15.** $23x+14y+11z=0$.

17. $4x+7y+2z+11=0$ **18.** $14x-5y+3z=16$.

19. $x^2+y^2+z^2-4x+12y+4z-5=0$; centre $(2, -1, -2)$, radius$=\sqrt{14}$.

20. $x^2+y^2+z^2-6x-4y+10z+12=0$.

Exercise 10 (*a*)

1. (*a*), (*d*), (*f*).
2. (*a*) 1 (*b*) 0 (*c*) 0 (*d*) 0.

Exercise 10 (*c*)

1. (*a*) It is hot or it is blowing.
(*b*) It is hot and it is blowing.
(*c*) It is hot but it is not blowing.
(*d*) It is not hot but it is blowing.

2. (*a*) $p+q$ (where $p=$He must stop ; $q=$He will faint).
(*b*) pq (where $p=$The sky is blue ; $q=$The grass is green).
(*c*) pq (where $p=$He is intelligent ; $q=$She is beautiful).
(*d*) $p+q$ (where $p=$Ice is cold ; $q=$10 is a prime).

3. (*i*) pq (*ii*) pq' (*iii*) $p'q'$ (*iv*) $p+p'q$ (*v*) $p'q'$ (*vi*) $p'+q$.

4. True for all statements p, q, r : (*a*).
False for all statements p, q, r : (*b*), (*d*), (*f*), (*g*).
Sometimes true and sometimes false : (*c*), (*e*).

Exercise 10 (*e*)

1. (*a*) $q \rightarrow p$ (*b*) $q \rightarrow p'$ (*c*) $p \rightarrow q$ (*d*) $p \rightarrow q$
(*e*) $p \rightarrow q$ (*f*) $q' \rightarrow p'$.
2. (*a*) 1 (*b*) 1 (*c*) 1 (*d*) 1 (*e*) 1 (*f*) 0.
3. If $\frac{1}{5}$ is greater than $\frac{1}{6}$, then 6 is greater than 5 ; If 6 is not greater than 5, then $\frac{1}{5}$ is not greater than $\frac{1}{6}$; If $\frac{1}{5}$ is not greater than $\frac{1}{6}$, then 6 is not greater than 5 ; 6 is greater than 5 but $\frac{1}{5}$ is not greater than $\frac{1}{6}$.

Exercise 10 (*f*)

3. No. 5. I didn't take tea in the evening today.

Exercise 10 (g)

2. $(x+y)(x'+y)(x'+y')$, $(x+y+z)(xyz+x+xy)$,
$xyza'+x'z+y'z+za$.
3. z, abd, $y'+z'$, $xy+xz+yz'+x'y'z$.

Test Your Understanding X

1. (*b*) 2. (*a*) 3. (*d*) 4. (*d*) 5. (*a*)
6. (*a*) 7. (*b*) 8. (*d*) 9. (*a*) 10. (*b*).

Review Exercise X

1. (*a*) Some roses are not red.
(*b*) It is not a rainy day.
(*c*) Some men are liars.
(*d*) $1+1 \neq 11$.
2. (*a*) Either Saurabh knows English or he is smart.
(*b*) Saurabh knows English and he is smart.
(*c*) Saurabh does not know English but he is smart.
(*d*) Saurabh neither knows English nor he is smart.

Exercise 11 (*a*)

1. {0, 1, 2,......, 500}. 2. {0, 1, ..., 10}.
3. {(D, D), (D, P), (P, D), (P, P)}, where (D, P) denotes that the first item is defective and the second one is perfect etc. Or {0, 1, 2} if we are noting the number of defective items.
4. {[W_1, P_1], [W_1, P_2], [W_2, P_1], [W_2, P_1], [W_1, W_2], [P_1, P_2]}, where square brackets indicate the result of the draw ; order immaterial.
5. Infinite. 6. {0}, {1}, {2}, ..., {10}.

7. $\{3, 4, 5, \ldots\ldots, 18\}$.

(a) $\{13, 14, \ldots\ldots, 18\}$. (b) $\{1, 2, \ldots\ldots, 5\}$.

8. (a) $\{0, 1, 2, \ldots\ldots, 99\}$ or $\{x : 0 \leqslant x \leqslant 99\}$;

(b) $\{51, 52, \ldots\ldots, 500\}$; (c) $\{x : 10 \leqslant x < 490\}$;

(d) $\{0\}$; (e) $\{0, 1, \ldots\ldots, 500\}$; (f) ϕ.

9. Yes, if the number of sprouts belongs to the set $\{51, 52, \ldots, 99\}$. No, the occurrence of either excludes the other.

10. $A=\{2, 4, 6\}$, $B=\{1, 3, 5\}$. Yes.

Exercise 11 (b)

1. $A \cup B$ = At the most one radio-set is defective ; $A \cap B=\phi$.

2. $S=\{1, 2, \ldots, 105\}$; $\sim E=\{53, 54, \ldots, 105\}$; $E \cap F=\{2, 4, 6, \ldots\ldots, 52\}$; $E \cup F=\{1, 2, 3, \ldots, 51, 52, 54, 56, \ldots, 102, 104\}$.

3. (a) Yes, because $A \cap B=\phi$.

(b) No, because $E \cup F \neq S$ in as much as $4 \notin S$.

4. $S=\{1, 2, 3, \ldots, 8\}$; $\sim A=\{1, 2, 3\}$: $\sim B=\{3, 4, \ldots, 8\}$;

$A \cap B=\phi$; $A \cup B=S \sim \{3\}$; $(\sim A) \cap B=B=\{1, 2\}$;

$(\sim A) \cup B=\sim A$. $A \cap B$ and each of the other events listed are mutually exclusive; also, $\sim B$ and $(\sim A) \cap B$ are mutually exclusive. $\sim A$ and $\sim B$ are exhaustive ; $\sim A$ and $A \cup B$ are exhaustive $\sim B$ and $A \cup B$ are exhaustive and so are $\sim B$ and B, as well as $\sim B$ and $(\sim A) \cup B$; $A \cup B$ and $(\sim A) \cup B$ are exhaustive.

Exercise 11 (c)

1. (a) $\frac{1}{6}$, (b) $\frac{1}{3}$, (c) $\frac{1}{2}$, (d) $\frac{1}{3}$,

2. (a) $\frac{1}{4}$, (b) $\frac{3}{4}$, (c) $\frac{1}{4}$.

3. (a) $\frac{3}{4}$, (b) $\frac{7}{12}$, (c) $\frac{1}{4}$.

4. $\frac{9}{13}$. 5. (a) $\frac{1}{6}$, (b) $\frac{1}{12}$.

6. $\frac{1}{6}$. 7. (a) $\frac{1}{55}$, (b) $\frac{1}{220}$, (c) $\frac{21}{55}$,

(d) $\frac{28}{55}$, (e) $\frac{41}{55}$, (f) $\frac{3}{11}$.

8. (a) $\frac{26}{477}$, (b) $\frac{2}{477}$, (c) $\frac{35}{477}$, (d) $\frac{8}{477}$.

9. 0·5. 10. 10 : 3. 11. (a) $p : 1-p$, (b) $1-p : p$.

Exercise 11 (*d*)

1. (*a*) 0·6, (*b*) 0·2, (*c*) 0·5, (*d*) 0·1.
2. (*a*) 0·7, (*b*) 0·6, (*c*) 0·7, (*d*) 0·3, (*e*) 0·3, (*f*) 0·4.
3. $\frac{5}{6}$. 4. $\frac{4}{13}$. 5. 0·4. 6. 0·39.
7. 1·2. 8. 0·95. 9. 0·6.
10. (*a*) 0·4, (*b*) 0·7.

Exercise 11 (*e*)

1. (*a*) $\frac{5}{14}$, (*b*) $\frac{5}{6}$, (*c*) $\frac{1}{4}$.
   Also, (*a*) is P (A | B) and (*b*) is P (B | A).
2. (*a*), (*b*), (*d*) 0·2, (*c*) 0·1. 3. 0·52 approx. 4. ·53 approx.
5. $\frac{4}{663}$. 6. $\frac{5}{221}$. 7. 0·21 approx.
8. (*a*) $\frac{66}{149}$, (*b*) $\frac{200}{447}$. 9. $\frac{1}{969}$. 10. $\frac{1}{7}$.

Exercise 11 (*f*)

1. (*i*) 0·12, (*ii*) 0·18, (*iii*) 0·58, (*iv*) 0·42.
2. (*a*) 0·01, (*b*) 0·01.
3. (*a*) 0·271, (*b*) 0·028, (*c*) 0·001, (*d*) 0·729.
4. P (A) $= \frac{1}{2}$ and P (B) $= \frac{1}{3}$ or else P (A) $= \frac{1}{3}$ and P (B) $= \frac{1}{2}$

   Obviously not.
5. Dinesh, Naresh $\frac{1}{2}$ each ; Suresh $\frac{3}{8}$. 6. 0·8645.
7. 0·0256. 9. 0·75. 10. (*i*) $\frac{13}{15}$, (*ii*) $\frac{2}{15}$.
11. $\frac{25}{56}$. 12. 0·525. 13. $\frac{19}{42}$. 14. $\frac{13}{32}$.
16. $\frac{37}{2^8}$. 17. $\frac{11}{2^6}$.

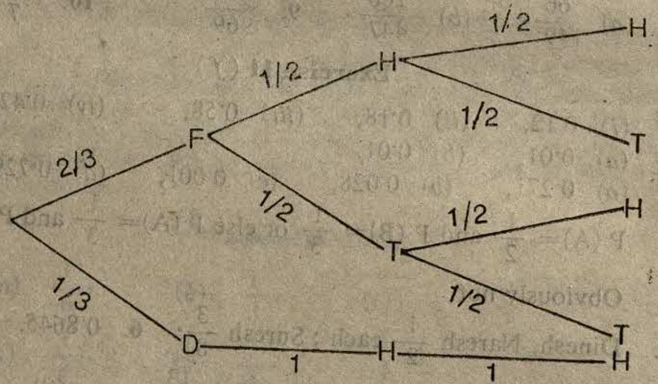
Test Your Understanding XI

1. (*d*). 2. (*d*). 3. (*c*). 4. (*d*). Every superset of a sample space is also a sample space. 5. (*d*). 6. (*a*).
7. (*b*). 8. (*c*). 9. (*c*). 10. (*b*). 11. (*c*).
12. (*c*). 13. (*a*). 14. (*c*). 15. (*c*). 16. (*d*).
17. (*a*). 18. (*i*) (*b*), (*ii*) (*b*).

Review Exercise XI

1. S_i is a sample point whereas $\{S_i\}$ is an elementary event.
4. $l/(l+m)$.
6. A is the event that the red die shows a 4 and B is the event that the number on each die is greater than 2. (*a*) $A \cap B$ is the event that the red die shows a 4 and the green die a number greater than 2. (*b*) $P(A)=\frac{1}{6}$. $P(B)=\frac{9}{4}$.

   $P(A \cap B)=\frac{1}{9}$. $P(A \cup B)=\frac{1}{2}$.
7. The data are insufficient.
8. (*a*) Independent. (*b*) Independent if the first marble is replaced before the second is drawn; not independent otherwise. (*c*) Not independent but pairwise independent.
9. (*a*) $\frac{1}{6}$, (*b*) $\frac{1}{2}$. (*c*) $\frac{1}{3}$.



Here F denotes the fair coin and D the one that has a head on both sides.

Yes. The events in (*a*), (*b*) and (*c*) are mutually exclusive and exhaustive. Hence the third probability can be obtained by subtracting the sum of the first two from 1.

10. $\frac{1}{3}$.
11. 0·01 (0·02, 0·03,......, 0·09, 0·1, 0·09,......, 0·01 respectively).
12. 0·08 approximately $\left(={}^{100}C_{50}\cdot\left(\frac{1}{2}\right)^{100}\right)$.
13. (*a*) May be; but not because the elements are listed in different orders. Order of elements in listing a set is immaterial.

(*b*) Yes. If S is one sample space, so are S ∪ T, T being any set. (*c*) No. ϕ (the impossible) and $\{a, b, c\}$ (the sure) events have not been listed. (*d*) The doctor is misleading the patient ; the probability of his dying from this disease is $\frac{9}{10}$ and not 0 ; the events of various patients dying from the disease in question being independent. (*e*) No and no ; the ratio of boys and girls will remain the same, no matter what the number of children, the tribe being a big one. (*f*) The probability of survival of the cancer patient from cancer is $\frac{1}{10}$. *Finally* every one must die.

14. (*i*) $\frac{1}{8}$. (*ii*) $\frac{7}{8}$. **15.** $\frac{2}{9}$. **16.** $\frac{5}{18}$.

17. (*a*) $\frac{188}{295}$, (*b*) $\frac{11}{295}$, (*c*) $\frac{48}{295}$.

18. $\frac{88\,!}{100\,!} \cdot \frac{90\,!}{78\,!}$. **19.** $\frac{95\,!}{100\,!} \cdot \frac{90\,!}{85\,!}$.

20. (*a*) The sample spaces on which the events are defined are different in the two cases. (It is like solving two quadratic equations in x, getting $x=1, 2$ say for one and $x=3, 4$ say for the other and then saying — how can 1, 2 be the same as 3, 4 ?) (*b*) the argument — *probability of getting at least one six in four throws is* $4 \cdot \frac{1}{6}$ — is fallacious. You can add the probabilities only if the events are mutually exclusive etc. (*c*) Same argument as in (*b*). The correct probability is $\cdot 491 < \frac{1}{2}$ and that is why De Mere's expectation was falsified in practice.

22. 25, the probability being ·506 approximately. Yes ; 25 or more.

Exercise 12 (*a*)

1.

X	0	1
f(X)	$\frac{1}{2}$	$\frac{1}{2}$

2.

Y	0	1	2
f(Y)	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

3.

X	0	1	2
f(X)	$\left(\frac{12}{13}\right)^2$	$\left(\frac{24}{13}\right)^2$	$\left(\frac{1}{13}\right)^2$

4.

X	0	1	2
f(X)	$\frac{2}{5}$	$\frac{8}{15}$	$\frac{1}{15}$

5.

X	0	1	2
f(X)	1–p–q+pq	p+q–2pq	pq

6.

X	1	2	3	...	14
f(X)	$\frac{39}{52}$	$\frac{13}{52}\cdot\frac{39}{51}$	$\frac{13}{52}\cdot\frac{12}{51}\cdot\frac{39}{50}$	...	$\frac{13}{52}\cdot\frac{12}{51}\cdot\frac{11}{50}\cdots\cdots\frac{2}{41}\cdot\frac{1}{40}\cdot 1$

Exercise 12 (*b*)

1.

X	0	1	2	3
f(X)	0.212	0.509	0.255	0.024

2. (*a*) 0·9688 ; 2·0587 ; 1·4348.
   (*b*) 0·5 ; 1·25 ; 1·118.
   (*c*) q ; pq ; pq.
3. 1·5, q.
4. 3·5 ; 2·92 approx.
5. To lose.
6. 112·5 paise.
7. 35 ; 300.
8. $\frac{1}{21}$, 4·3, 2·2.
9. The customer stands to lose on an average Rs. 5·26 per bet in the long run.

10.

X	0	1	3
f(X)	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

; E (X)=1.

Exercise 12 (*c*)

1. $\frac{57}{64}$. 2. (*a*) $\left(\frac{5}{6}\right)^6$, (*b*) ·0087. 3. 0·9375. 4. 0·784.

5. $^{400}C_4\left(\frac{1}{36}\right)^4\left(\frac{35}{36}\right)^{396}$.

6. (*a*)

X	0	1	2	3	4	5	6
f(X)	0.0878	0.2634	0.3292	0.2195	0.0823	0.0165	0.0014

(*b*)

X	0	1	2	3	4
f(X)	$\left(\frac{5}{6}\right)^4$	$\frac{5}{9}\left(\frac{5}{6}\right)^2$	$\frac{1}{6}\left(\frac{5}{6}\right)^2$	$\frac{5}{9}\left(\frac{1}{6}\right)^2$	$\left(\frac{1}{6}\right)^4$

7. (*a*) 0·3292, (*b*) 0·0823, (*c*) 0·0878.
9. $1-23\times2^{-22}$.
10. (*i*) $(\cdot97)^{10}$, (*ii*) $3\,(\cdot97)^{99}$, (*iii*) $4950\,(\cdot03)^2\,(\cdot97)^{98}$.
11. 0·5.
12.

X	0	1	2	3	4	5	6	7	8	9
f(X)	.0260	.1171	.2341	.2731	.2048	.1024	.0341	.0073	.0009	.0001

13. Mean=1000 ; profit=Rs. 2000 > advertisement cost. Hence advisable.
14. 1 ; 0·866.
15. Yes ; student made a mistake somewhere, because variance 16 > 12, the mean. It should be the other way. 4 and 12 cannot be the mean and S.D. of a binomial distribution in any order. [Why ?]

Test Your Understanding XII

1. (*b*). 2. (*d*). 3. (*d*). 4. (*d*). 5. (*c*). 6. (*b*).
7. (*c*). 8. (*d*). 9. (*c*). 10. (*d*). 11. (*c*).

Review Exercise XII

1.

X	1	2	3
f(X)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

2.

X	1	2	3	4
f(X)	$\frac{2}{5}$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{10}$

4. $\frac{n}{12}, \frac{n}{12}(n+2)$.

5.

X	1	2	3	.....	n	....
f(X)	$\frac{1}{2}$	$\left(\frac{1}{2}\right)^2$	$\left(\frac{1}{2}\right)^3$	.....	$\left(\frac{1}{2}\right)^n$	....

Yes.

8. No ; trials are not independent. Probability of a day being a rainy day increases if this day follows a rainy day.
9. 0·1692, approx. 11. 3. 12. 0·5 ; $n/2^{n-1}$.
13. (*i*) 0·2461, (*ii*) 0·4102.

Exercise 13 (*a*)

2.

Y \ X	1	2	3	4	Totals
1	1	1	3	–	5
2	1	2	–	–	3
3	1	–	2	1	4
4	–	–	3	–	3
Totals	3	3	8	1	15

3.

X	1	2	3	4
f_X	3	3	8	1

Y	1	2	3	4
f_Y	5	3	4	3

4.

X	5	6	7	8
f_X	3	9	7	11

Y	3	4	5
f_Y	5	9	16

The aptitude index of the employees for their job and their output respectively.

5. (*a*)

X	1	2	3	4
$\bar{Y}$	2	1.67	2.63	3

Y	1	2	3	4
$\bar{X}$	2.4	1.67	2.75	1.13

(*b*)

X	5	6	7	8
$\bar{Y}$	3.67	3.78	4.71	4.81

Y	3	4	5
$\bar{X}$	5.8	6.44	7.43

6. (*a*)

Y \ X	4	6	7	8	9	10	12	13	15	16	17	Totals
2	1	–	–	–	–	–	–	–	–	–	–	1
6	–	1	–	–	–	–	–	–	–	–	–	1
7	–	–	1	–	–	–	–	–	–	–	–	1
8	–	–	–	1	1	–	–	–	–	–	–	2
10	–	–	–	–	–	1	1	–	–	–	–	2
11	–	–	–	–	–	–	–	1	1	–	1	3
12	–	–	–	–	–	1	–	1	–	1	–	3
15	–	–	–	–	–	–	–	–	–	–	1	1
16	–	–	–	–	–	–	–	–	–	1	–	1
Totals	1	1	1	1	1	2	1	2	1	2	2	15

(*b*)

X	4	6	7	8	9	10	12	13	15	16	17
f_X	1	1	1	1	1	2	1	2	1	2	2

Y	2	6	7	8	10	11	12	15	16
fY	1	1	1	2	2	3	3	1	1

(*c*)

X	4	6	7	8	9	10	12	13	15	16	17
$\overline{Y}$	2	6	7	8	8	11	10	11.5	11	14	13

Y	2	6	7	8	10	11	12	15	16
$\bar{X}$	4	6	7	8.5	11	15	13	17	16

(*e*) Yes ; the dots form sort of a band. Several straight lines can be drawn close to most of the points because *close* is quite vague. Wait till you reach the last section of the chapter ; *close* would not be so vague then.

7. $\bar{X}=11{\cdot}53$, $\bar{Y}=10{\cdot}06$, $\sigma_X=4{\cdot}07$, $\sigma_Y=3{\cdot}42$. No, the dots are still close to a line. All lines close to these dots have nearly the same slope which is the same as that for exercise 6 above.

Exercise 13 (*b*)

1. (*a*) –, (*b*) –, (*c*) +, (*d*) +, (*e*) –, (*f*) +, (*g*) 0, as per specific evidence.
2. No. 3. Positive.
4. Perfect positive correlation.
5. Positive correlation. 6. –·33 ; ·98 ; 1 ; ·96.
7. ·69. 8. ·52.
9. ·95. 10. ·44.
11. –·76. 14. ·21.
15. ·75.

Test Your Understanding XIII

1. (*d*). 2. (*d*). 3. (*b*). 4. (*c*). 5. (*c*).
6. (*d*). 7. (*a*). 8. (*c*), for all the parts. 9. (*c*).
10. (*c*). 11. (*c*). 12. (*d*). 13. (*d*).
14. (*d*). 15. (*c*).

Review Exercise XIII

1. ·98. 2. ·81. 3. ·69. 4. 0·92.
5. 0·78. 6. 0·4. Take $N=40$ and $\Sigma x^2=236$.
7. (*a*) Yes ; curvilinear. (*b*) Considerably smaller
   (*c*) No. (*d*) Yes. (*e*) (*i*) –·06, (*ii*) –·91, (*iii*) ·85.
   [*r* would be small even for strong relationship unless the relationship is linear.]
10. (*a*) Yes.
12. (*a*) False (F). The pairs of values must relate to the same unit of observation.
    (*b*) F. r will be low when the relationship is curvilinear.

(*c*) F. It would be zero.

(*d*) Not *curvilinear* but *linear*.

(*e*) F. Larger values of X are associated with smaller values of Y.

(*f*) Not necessarily.

(*g*) *Higher*, not lower.

(*h*) False.

(*i*) F. If gives the mean of the various values of Y corresponding to the given value of X.

(*j*) F. It is used to predict Y from X.

(*k*) True.

Exercise 14 (*a*)

1. All are false.
2. 10, 100 1000, 10000, 100000, 111, 1010, 1111, 10100.
3. 5, 13, 21, 15, 14.
4. (*i*) 1111, (*ii*) 10111, (*iii*) 10100.
5. 3. 7. 15. 4.
6. Information is processed data.

Exercise 14 (*c*)

1. (*c*) and (*d*). 2. (*b*), (*c*) and (*d*).

3. (*a*) to (*d*) at the time of problem-analysis ;
   (*e*) while writing the algorithm and (*f*) while looking for a method of solution.

4. ({(SP, CP), SP−CP}, {A_1, A_2, A_2', A_2''} where $A_1 \leftrightarrow$ (SP−CP) ; $A_2 \leftrightarrow$ (loss is CP−SP) when SP−CP $<$ 0, $A'_2 \leftrightarrow$ (no loss no gain) when SP−CP=0, $A''_2 \leftrightarrow$ (profit is SP−CP) if SP−CP$>$0. Answer not unique.

7. Calculates the maximum M of N given number R[1], ..., R [N] and tells which of these (Jth) is the maximum.

8. (*a*) A and B would be interchanged.
   (*b*) Both A and B have the same value as B.

Test Your Understanding XIV

1. (*d*). 2. (*b*). 3. (*c*). 4. (*b*). 5. (*b*).
6. (*c*). 7. (*b*). 8. (*c*). 9. (*c*). 10. (*d*).

Exercise 15 (*a*)

1. (*a*) 3·7961. (*b*) 2·00095. (*c*) ·0079. (*d*) −·151.
   (*e*) −33·149. (*f*) −5496.

2. (*a*) 2. (*b*) 4. (*c*) −2. (*d*) 1.
   (*e*) −3. (*f*) −4.

3. (*a*) 379·61. (*b*) 20009·5. (*c*) ·000079. (*d*) −1·51.
(*e*) −·033149. (*f*) −·5496.

4. (*a*) ·37426E4. (*b*) ·374265E 4 (*c*) 1·2345E4.
(*d*) ·00012345E4. (*e*) −·0012345E4.
(*f*) ·12345E4.

5. (*a*) 3742600E−3 (*b*) 3742650E−3.
(*c*) 12345000E−3. (*d*) 1234·5E−3.
(*e*) —12345E−3. (*f*) 1234500E−3.

6. (*a*) 12·34E−3. (*b*) 12·34E−2.
(*c*) 12·34E1. (*d*) 12·34E0.
(*e*) 12·34E1. (*f*) 12·34E2.
(*g*) 12·34E3. (*h*) 12·34E4.

7. (*a*) ·372E8. (*b*) ·372E5. (*c*) ·3765E−3. (*d*) ·999E10.
(*e*) ·1234E4. (*f*) ·57E−2. (*g*) ·61E 1. (*h*) ·2345E0.
(*i*) ·1E—3.

8. (*a*)

+	2	3	4	5	+	1	2

(*b*)

+	2	3	4	5	+	1	3

(*c*)

+	2	3	4	5	−	1	0

(*d*)

+	2	5	9	1	+	1	2

(*e*)

−	3	5	9	7	+	0	1

(*f*)

+	2	9	2	0	+	1	0

259·11E+09 and −359·79E−2, the respective errors being ·1E 8($=10^7$) and −·9E−3 (=−·009) or ·004% and ·09% upto three places of decimal.

Exercise 15 (*b*)

1. (*a*) ·9999E10. (*b*) ·8888E19.
(*c*) ·6678E5. (*d*) ·8899E4.
(*e*) ·3665E−5. (*f*) ·8772E−4.

(*g*) ·176E−5. (*h*) ·7374E−6.
(*i*) ·9887E1. (*j*) ·6710E2.
(*k*) ·8808E3. (*l*) ·9801E1.
(*m*) ·1429E8. (*n*) ·159E6.
(*o*) ·1079E95. (*p*) −·8605E7.

2. (*c*) ·66781E5. (*d*) ·889911E4.
(*g*) ·17608E−5. (*h*) ·737473E−6.
(*i*) ·988765E1. (*j*) ·6710365E2.
(*k*) ·88089996E3. (*l*) ·9801007342E1.
(*m*) ·14294E8. (*n*) 15904E6.
(*o*) ·107912E95.

3. (*d*) and (*f*).

Exercise 15 (*d*)

1. (*a*) ·628E6. (*b*) ·127E8. (*c*) ·679E−1.
(*d*) Overflow. (*e*) Underflow. (*f*) ·273E0.
2. All would change. (*a*) ·62868E+06.

Exercise 15 (*e*)

1. 2347, 2457, 1120, 1256, 1246, 1232, 1263, 1263, 1266, 1266, 1260, 4270, 4268.
2. 3·798, 4·568, 2·091, 2·91, 2·998, 2·666, ·746, ·846, ·667, ·555.
3. Correct.

Exercise 15 (*f*)

1. ·00982, ·00451, ·002956, ·0076.
2. ·0004, ·0002. ·0005, ·0005. For 3·7984 and 7·9125.
3. 35800, 26400, 31000, 4000.
(*a*) −87, −1, −9, −2. (*b*) 87, 1, 9, 2.
(*c*) Rounded off to 3 decimal places ·243E−2, ·379E−4, ·29E−3, ·5E−3.
(*d*) ·243, ·379E−2, ·29E−1, ·5E−1.

Exercise 15 (*g*)

1. (*a*) −·9. (*b*) −·3. (*c*) −9·18. (*d*) 0. (*e*) ·3.
2. ·006 for both parts. 3. ·0501 for both.
4. With an error bound ·5 cm for each.

Exercise 15 (*h*)

1. (*a*) ·5265E2, ·5264E2. True value 52·654. The first.
(*b*) ·3841E4, ·3842E4. True value 3841·107. The first,
2. ·416E−2, ·4166E−2. The latter.

3. 9 ; 9·991. True value 8·9991. The first output is less in error (because fewer operations were performed).
4. ·08, ·0796. The latter.
5. (*a*) ·56E−1, ·5574E−1. (*b*) ·4E−1, ·4174E−1.
(*c*) ·45E−2, ·4509E−2. (*d*) ·1E−1.
6. Inherent.

Exercise 15 (*i*)

1. (*a*) 7·39. (*b*) 12·2. (*c*) 24·5. (*d*) 4·06.
2. (*a*) 0·841. (*b*) −0·416. (*c*) 0·863. (*d*) 0·943.
(*e*) Same as for sin 3.
3. (*a*) 0·650. (*b*) −0·506. (*c*) 0·154. (*d*) −0·858.
4. (*a*) ·1823. (*b*) ·0296. (*c*) −·6931. (*d*) ·6931.
5. No ; 1·0986.

Exercise 15 (*j*)

1. (*a*) 1 and 2. (*b*) 0 and 1 ; 1 and 2. (*c*) 2 and 3.
3. 1·3. 4. (*a*) 2·09. (*b*) 2·38. (*c*) 2·96.
(*d*) 2·16. 5. ·3984.

Exercise 15 (*k*)

1. (*a*) 2·706. (*b*) 2·620. (*c*) 1·731. (*d*) 1·324.
2. 1·406. 3. 1·7. 4. 1·05. 5. ·588.

Exercise 15 (*l*)

1. (*a*) 1·813. (*b*) 1·817. (*c*) 1·856. (*d*) 4·264.
(*e*) 2·104. (*f*) 1·325.
3. 1·7320, 2·2360, 2·6458. 4. 2·7984. 5. ·511.

Exercise 15 (*m*)

1. (*a*), (*b*) and (*c*). 2. Interchange first and second equations for (*a*), and first and third for (*b*).
3. (*a*) $x_1=1,\ x_2=2.$ (*b*) $x=\cdot79,\ y=\cdot70.$
(*c*) $x_1=1,\ x_2=2,\ x_3=-1.$
(*d*) $x_1=2\cdot56,\ x_2=1\cdot72,\ x_3=-1\cdot06.$
4. Magnitude of the coefficient of the pivoted variable should be greater than or equal to the sum of the magnitudes of the coefficients of the remaining variables, with at least one inequality being strict.
5. $x_1=46\cdot15,\ x_2=84\cdot61,\ x_3=92\cdot31,\ x_4=84\cdot61,\ x_5=46\cdot15.$
6. Lesser iterations are required.

Exercise 15 (*n*)

1. (*a*) 0·696. (*b*) 0·782. (*c*) 0·771. (*d*) 1·14.
The actual values are respectively 0·693, 0·7854, $\pi/4$ and 1·0.
2. (*a*) 1·7683, (*c*) 1·7728, (*c*) 1·7904. The respective errors are −0·0013, −0·0058, −0·0234. The first.
3. 23·9944, −0·08. 4. (*a*) 0·708, (*b*) 0·697, (*c*) 0·694. Actual value 0·693 to three decimal places. Errors 0·015, 0·004 and 0·001.
5. (*a*) 0·694, (*b*) 0·693, (*c*) 0·693. Simpson's rule; it converges earlier.

Test Your Understanding XV

1. (*b*).	2. (*b*).	3. (*c*).	4. (*b*).
5. (*a*).	6. (*d*).	7. (*c*).	8. (*c*).
9. (*a*).	10. (*b*).	11. (*c*).	12. (*d*).
13. (*a*).	14. (*b*).	15. (*a*).	16. (*b*).

Review Exercise XV

1. −2. 2. (*a*)

+	2	5	9	+	0	2

(*b*)

−	2	5	3	+	0	2

(*c*)

+	1	5	1	+	0	3

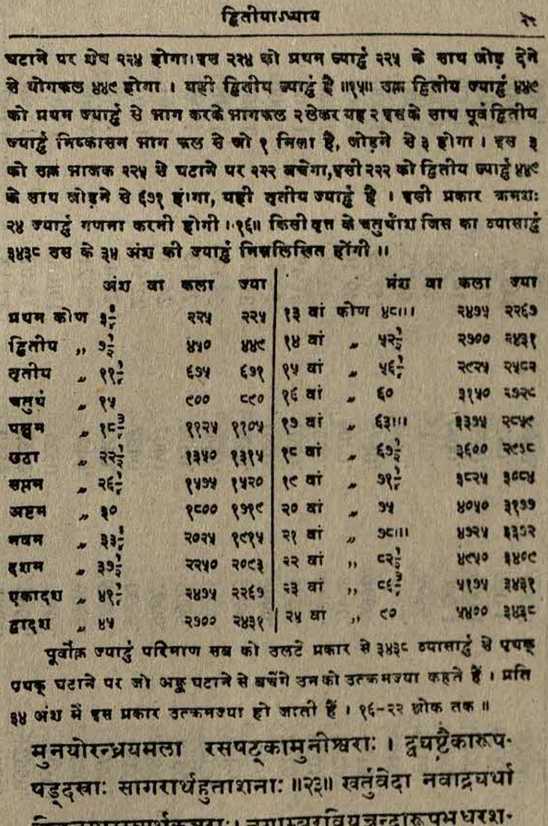
(*d*)

+	1	0	0	−	0	2

4. Inherent. 5. 13·42. 6. 2·72. 7. [−3, −2], [0, 1] and [2, 3]. 8. 1·4142· 9. 1·9999.
10. 0·6071. 12. 177·483.
13. Rounded off to two decimal places, $x=2{\cdot}12$, $y=1{\cdot}4$, $z=3{\cdot}06$. 15. Only (*c*).

TABLES AND APPENDICES

★ ★ ★



द्वितीयाऽध्याय ५९

घटाने पर शेष २२४ होगा। इस २२४ को प्रथम ज्यार्द्ध २२५ के साथ जोड़ देने से योगफल ४४९ होगा । यही द्वितीय ज्यार्द्ध है ॥१५॥ उक्त द्वितीय ज्यार्द्ध ४४९ को प्रथम ज्यार्द्ध से भाग करके भागफल २ लेकर यह २ इसके साथ पूर्व द्वितीय ज्यार्द्ध मिष्कासन भाग फल से जो १ मिला है, जोड़ने से ३ होगा। इस ३ को उक्त भाजक २२५ से घटाने पर २२२ बचेगा, इसी २२२ को द्वितीय ज्यार्द्ध ४४९ के साथ जोड़ने से ६७१ होगा, यही तृतीय ज्यार्द्ध है । इसी प्रकार क्रमशः २४ ज्यार्द्ध गणना करनी होगी ।।१६॥ किसी वृत्त के चतुर्थांश जिस का व्यासार्द्ध ३४३८ उस के ३४ अंश की ज्यार्द्ध निम्नलिखित होंगी ॥

	अंश	वा कला	ज्या		अंश	वा कला	ज्या
प्रथम कोण	३ ३/४	२२५	२२५	१३ वां कोण	४८।।।	२४७५	२२६७
द्वितीय ,,	७ १/२	४५०	४४९	१४ वां ,,	५२ १/२	२७००	२४३१
तृतीय ,,	११ १/४	६७५	६७१	१५ वां ,,	५६ १/४	२९२५	२५८२
चतुर्थ ,,	१५	९००	८९०	१६ वां ,,	६०	३१५०	२७२८
पञ्चम ,,	१८ ३/४	११२५	११०५	१७ वां ,,	६३।।।	३३७५	२८५९
छठा ,,	२२ १/२	१३५०	१३१५	१८ वां ,,	६७ १/२	३६००	२९७८
सप्तम ,,	२६ १/४	१५७५	१५२०	१९ वां ,,	७१ १/४	३८२५	३०८४
अष्टम ,,	३०	१८००	१७१९	२० वां ,,	७५	४०५०	३१७७
नवम ,,	३३ ३/४	२०२५	१९१५	२१ वां ,,	७८।।।	४७२५	३३७२
दशम ,,	३७ १/२	२२५०	२०९३	२२ वां ,,	८२ १/२	४९५०	३४०९
एकादश ,,	४१ १/४	२४७५	२२६७	२३ वां ,,	८६ ३/४	५१७५	३४३१
द्वादश ,,	४५	२७००	२४३१	२४ वां ,,	९०	५४००	३४३८

पूर्वोक्त ज्यार्द्ध परिमाण सब को उलटे प्रकार से ३४३८ व्यासार्द्ध से पृथक् पृथक् घटाने पर जो अङ्क घटाने से बचेंगे उन को उत्क्रमज्या कहते हैं । प्रति ३४ अंश में इस प्रकार उत्क्रमज्या हो जाती हैं । १६-२२ श्लोक तक ॥

मुनयोरन्ध्रयमला रसषट्कामुनीश्वराः । द्व्यष्टैकारूपषड्दस्राः सागरार्थहुताशनाः ॥२३॥ खर्तुवेदा नवाद्रयर्था दिङ्नगास्त्र्यर्थकुञ्जराः। नगाम्बराबियच्चन्द्रारूपभूधरश-

FROM THE FIRST PRINTED EDITION OF THE SURYA SIDDHANTA
Printed at Meerut, India, c. 1867. This is the oldest Hindu work on astronomy

The earliest trigonometric tables are the ones in Surya Siddhanta. A page from Surya Siddhanta giving values of sines is reproduced above. The values given above are remarkably close to the modern values.

TABLE 1

Values of Trigonometric Functions

Angle	sin	cos	tan	cot	sec	csc	Angle
0°00′	·0000	1·000	·0000	—	1·000	—	90°00′
10′	·0029	1·000	·0029	343·8	1·000	343·8	50′
20′	·0058	1·000	·0058	171·9	1·000	171·9	40′
30′	·0087	1·000	·0087	114·6	1·000	114·6	30′
40′	·0116	·9999	·0116	85·94	1·000	85·95	20′
50′	·0145	·9999	·0145	68·75	1·000	68·76	10′
1°00′	·0175	·9998	·0175	57·29	1·000	57·30	89°00′
10′	·0204	·9998	·0204	49·10	1·000	49·11	50′
20′	·0233	·9997	·0233	42·96	1·000	42·98	40′
30′	·0262	·9997	·0262	38·19	1·000	38·20	30′
40′	·0291	·9996	·0291	34·37	1·000	34·38	20′
50′	·0320	·9995	·0320	31·24	1·001	31·26	10′
2°00′	·0349	·9994	·0349	28·64	1·001	28·65	88°00′
10′	·0378	·9993	·0378	26·43	1·001	26·45	50′
20′	·0407	·9992	·0407	24·54	1·001	24·56	40′
30′	·0436	·9990	·0437	22·90	1·001	22·93	30′
40′	·0465	·9989	·0466	21·47	1·001	21·49	20′
50′	·0494	·9989	·0495	20·21	1·001	20·23	10′
3°00′	·0523	·9986	·0524	19·08	1·001	19·11	87°00′
10′	·0552	·9985	·0553	18·07	1·002	18·10	50′
20′	·0581	·9983	·0582	17·17	1·002	17·20	40′
30′	·0610	·9981	·0612	16·35	1·002	16·38	30′
40′	·0640	·9980	·0641	15·60	1·002	15·61	20′
50′	·0669	·9978	·0670	14·92	1·002	14·96	10′
Angle	**cos**	**sin**	**cot**	**tan**	**csc**	**sec**	**Angle**

Values of Trigonometric Functions

(Contd.)

Angle	sin	cos	tan	cot	sec	csc	Angle
4°00′	·0698	·9976	·0699	14·30	1·002	14·34	86°00′
10′	·0727	·9974	·0729	13·73	1·003	13·76	50′
20′	·0756	·9971	·0758	13·20	1·003	13·23	40′
30′	·0785	·9969	·0787	12 71	1·003	12·75	30′
40′	·0814	·9967	·0816	12·25	1·003	12·29	20′
50′	·0843	·9964	·0846	11·83	1·004	11·87	10′
5°00′	·0872	·9962	·0875	11·43	1·004	11·47	85°00′
10′	·0901	·9959	·0904	11·06	1·004	11·10	50′
20′	·0929	·9957	·0934	10·71	1·004	10·76	40′
30′	·0958	·9954	·0963	10·39	1·005	10·43	30′
40′	·0987	·9951	·0992	10·08	1·005	10·13	20′
50′	·1016	·9948	·1022	9·788	1·005	9·839	10′
6°00′	·1045	·9945	·1051	9·514	1·006	9·567	84°00′
10′	·1074	·9942	·1080	9·225	1·006	9·309	50′
20′	·1103	·9939	·1110	9·010	1·006	9·065	40′
30′	·1132	·9936	·1139	8·777	1·006	8·834	30′
40′	·1161	·9932	·1169	8·556	1·007	8·614	20′
50′	·1190	·9929	·1198	8·345	1·007	8·405	10′
7°00′	·1219	·9925	·1228	8·144	1·008	8·206	83°00′
10′	·1248	·9922	·1257	7·953	1·008	8·016	50′
20′	·1276	·9918	·1287	7·770	1·008	7·834	40′
30′	·1305	·9914	·1317	7·596	1·009	7·661	30′
40′	·1334	·9911	·1346	7·429	1·009	7·496	20′
50′	·1363	·9907	·1376	7·269	1·009	7·337	10′
Angle	cos	sin	cot	tan	csc	sec	Angle

Values of Trigonometric Functions (*Contd.*)

Angle	sin	cos	tan	cot	sec	csc	Angle
8°00′	·1392	·9903	·1405	7·115	1·010	7·185	82°00′
10′	·1421	·9899	·1435	6·968	1·010	7·040	50′
20′	·1449	·9894	·1465	6·827	1·011	6·900	40′
30′	·1478	·9890	·1495	6·691	1·011	6·765	30′
40′	·1507	·9886	·1524	6·561	1·012	6·636	20′
50′	·1536	·9881	·1554	6·435	1·012	6·512	10′
9°00′	·1564	·9877	·1584	6·314	1·012	6·392	81°00′
10′	·1593	·9872	·1614	6·197	1·013	6·277	50′
20′	·1622	·9868	·1644	6·084	1·013	6·166	40′
30′	·1650	·9863	·1673	5·976	1·014	6·059	30′
40′	·1679	·9858	·1703	5·871	1·014	5·955	20′
50′	·1708	·9853	·1733	5·769	1·015	5·855	10′
10°00′	·1736	·9848	·1763	5·671	1·015	5·759	80°00′
10′	.1765	·9843	·1793	5·576	1·016	5·665	50′
20′	.1794	·9838	·1823	5·485	1·016	5·575	40′
30′	.1822	·9833	·1853	5·396	1·017	5·487	30′
40′	.1851	·9827	·1883	5·309	1·018	5·403	20′
50′	.1880	·9822	·1914	5·226	1·018	5·320	10′
11°00′	·1908	·9816	·1944	5·145	1·019	5·241	79°00′
10′	·1937	·9811	·1974	5·066	1·019	5·164	50′
20′	·1965	·9805	·2004	4·989	1·020	5·089	40′
30′	·1994	·9799	·2035	4·915	1·020	5·016	30′
40′	·2022	·9793	·2065	4·843	1·021	4·945	20′
50′	·2051	·9787	·2095	4·773	1·022	4·876	10′
Angle	cos	sin	cot	tan	csc	sec	Angle

Values of Trigonometric Functions *(Contd.)*

Angle	sin	cos	tan	cot	sec	csc	Angle
12°00′	·2079	·9781	·2126	4·705	1·022	4·810	78°00′
10′	·2108	·9775	·2156	4·638	1·023	4·745	50′
20′	·2136	·9769	·2186	4·574	1·024	4·682	40′
30′	·2164	·9763	·2217	4·511	1·024	4·620	30′
40′	·2193	·9757	·2247	4·449	1·025	4·560	20′
50′	·2221	·9750	·2278	4·390	1·026	4·502	10′
13°00′	·2250	·9744	·2309	4·331	1·026	4·445	77°00′
10′	·2278	·9737	·2339	4·275	1·027	4·390	50′
20′	·2306	·9730	·2370	4·219	1·028	4·336	40′
30′	·2334	·9724	·2401	4·165	1·028	4·284	30′
40′	·2363	·9717	·2432	4·113	1·029	4·232	20′
50′	·2391	·9710	·2462	4·061	1·030	4·182	10′
14°00′	·2419	·9703	·2493	4·011	1·031	4·134	76°00′
10′	·2447	·9696	·2524	3·962	1·031	4·086	50′
20′	·2476	·9689	·2555	3·914	1·032	4·039	40′
30′	·2504	·9681	·2586	3·867	1·033	3·994	30′
40′	·2532	·9674	·2617	3·821	1·034	3·950	20′
50′	·2560	·9667	·2648	3·776	1·034	3·906	10′
15°00′	·2588	·9659	·2679	3·732	1·035	3·864	75°00′
10′	·2616	·9652	·2711	3·689	1·036	3·822	50′
20′	·2644	·9644	·2742	3·647	1·037	3·782	40′
30′	·2672	·9636	·2773	3·606	1·038	3·742	30′
40′	·2700	·9628	·2805	3·566	1·039	3·703	20′
50′	·2728	·9621	·2836	3·526	1·039	3·665	10′
Angle	**cos**	**sin**	**cot**	**tan**	**csc**	**sec**	**Angle**

Values of Trigonometric Functions (*Contd.*)

Angle	sin	cos	tan	cot	sec	csc	Angle
16°00′	·2756	·9613	·2867	3·487	1·040	3·628	74°00′
10′	·2784	·9605	·2899	3·450	1·041	3·592	50′
20′	·2812	·9596	·2931	3·412	1·041	3·556	40′
30′	·2840	·9588	·2962	3·376	1·043	3·521	30′
40′	·2868	·9580	·2994	3·340	1·044	3·487	20′
50′	·2896	·9572	·3026	3·305	1·045	3·453	10′
17°00′	·2924	·9563	·3057	3·271	1·046	3·420	73°00′
10′	·2952	·9555	·3089	3·237	1·047	3·388	50′
20′	·2979	·9546	·3121	3·204	1·048	3·356	40′
30′	·3007	·9537	·3153	3·172	1·049	3·326	30′
40′	·3035	·9528	·3185	3·140	1·049	3·295	20′
50′	·3062	·9520	·3217	3·108	1·050	3·265	10′
18°00′	·3090	·9511	·3249	3·078	1·051	3·236	72°00′
10′	·3118	·9502	·3281	3·047	1·052	3·207	50′
20′	·3145	·9492	·3314	3·018	1·053	3·179	40′
30′	·3173	·9483	·3346	2·989	1·054	3·152	30′
40′	·3201	·9474	·3378	2·960	1·056	3·124	20′
50′	·3228	·9465	·3411	2·932	1·057	3·098	10′
19°00′	·3256	·9455	·3443	2·904	1·058	3·072	71°00′
10′	·3283	·9446	·3476	2·877	1·059	3·046	50′
20′	·3311	·9436	·3508	2·850	1·060	3·021	40′
30′	·3338	·9426	·3541	2·824	1·061	2·996	30′
40′	·3365	·9417	·3574	2·798	1·062	2·971	20′
50′	·3393	·9407	·3607	2·773	1·063	2·947	10′
Angle	cos	sin	cot	tan	csc	sec	Angle

Values of Trigonometric Functions (*Contd.*)

Angle	sin	cos	tan	cot	sec	csc	Angle
20°00′	·3420	·9397	·3640	2·747	1·064	2·924	70°00′
10′	·3448	·9387	·3673	2·723	1·065	2·901	50′
20′	·3475	·9377	·3706	2·699	1·066	2·878	40′
30′	·3502	·9367	·3739	2·675	1·068	2·855	30′
40′	·3529	·9356	·3772	2·651	1·069	2·833	20′
50′	·3557	·9346	·3805	2·628	1·070	2·812	10′
21°00′	·3584	·9336	·3839	2·605	1·071	2·790	69°00′
10′	·3611	·9325	·3872	2·583	1·072	2·769	50′
20′	·3638	·9315	·3906	2·560	1·074	2·749	40′
30′	·3665	·9304	·3939	2·539	1·075	2·729	30′
40′	·3692	·9293	·3973	2·517	1·076	2·709	20′
50′	·3719	·9283	·4006	2·496	1·077	2·689	10′
22°00′	·3746	·9272	·4040	2·475	1·079	2·669	68°00′
10′	·3773	·9261	·4074	2·455	1·080	2·650	50′
20′	·3800	·9250	·4108	2·434	1·081	2·632	40′
30′	·3827	·9239	·4142	2·414	1·082	2·613	30′
40′	·3854	·9228	·4176	2·394	1·084	2·595	20′
50′	·3881	·9216	·4210	2·375	1·085	2·577	10′
23°00′	·3907	·9205	·4245	2·356	1·086	2·559	67°00′
10′	·3934	·9194	·4279	2·337	1·088	2·542	50′
20′	·3961	·9182	·4314	2·318	1·089	2·525	40′
30′	·3987	·9171	·4348	2·300	1·090	2·508	30′
40′	·4014	·9159	·4383	2·282	1·092	2·491	20′
50′	·4041	·9147	·4417	2·264	1·093	2·475	10′
Angle	cos	sin	cot	tan	csc	sec	Angle

Values of Trigonometric Functions

(*Contd.*)

Angle	sin	cos	tan	cot	sec	csc	Angle
24°00′	·4067	·9135	·4452	2·246	1·095	2·459	66°00′
10′	·4094	·9124	·4487	2·229	1·096	2·443	50′
20′	·4120	·9112	·4522	2·211	1·097	2·427	40′
30′	·4147	·9100	·4557	2·194	1·099	2·411	30′
40′	·4173	·9088	·4592	2·177	1·100	2·396	20′
50′	·4200	·9075	·4628	2·161	1·102	2·381	10′
25°00′	·4226	·9063	·4663	2·145	1·103	2·366	65°00′
10′	·4253	·9051	·4699	2·128	1·105	2·352	50′
20′	·4279	·9038	·4734	2·112	1·106	2·337	40′
30′	·4305	·9026	·4770	2·097	1·108	2·323	30′
40′	·4331	·9013	·4806	2·081	1·109	2·309	20′
50′	·4358	·9001	·4841	2·066	1·111	2·295	10′
26°00′	·4384	·8988	·4877	2·050	1·113	2·281	64°00′
10′	·4410	·8975	·4913	2·035	1·114	2·268	50′
20′	·4436	·8962	·4950	2·020	1·116	2·254	40′
30′	·4462	·8949	·4986	2·006	1·117	2·241	30′
40′	·4488	·8936	·5022	1·991	1·119	2·228	20′
50′	·4514	·8923	·5059	1·977	1·121	2·215	10′
27°00′	·4540	·8910	·5095	1·963	1·122	2·203	63°00′
10′	·4566	·8897	·5132	1·949	1·124	2·190	50′
20′	·4592	·8884	·5169	1·935	1·126	2·178	40′
30′	·4617	·8870	·5206	1·921	1·127	2·166	30′
40′	·4643	·8857	·5243	1·907	1·129	2·154	20′
50′	·4669	·8843	·5280	1·894	1·131	2·142	10′
Angle	cos	sin	cot	tan	csc	sec	Angle

Values of Trigonometric Functions (*Contd.*)

Angle	sin	cos	tan	cot	sec	csc	Angle
28°00′	·4695	·8829	·5317	1·881	1·133	2·130	62°00′
10′	·4720	·8816	·5354	1·868	1·134	2·118	50′
20′	·4746	·8802	·5392	1·855	1·136	2·107	40′
30′	·4772	·8788	·5430	1·842	1·138	2·096	30′
40′	·4797	·8774	·5467	1·829	1·140	2·085	20′
50′	·4823	·8760	·5505	1·816	1·142	2·074	10′
29°00′	·4848	·8746	·5543	1·804	1·143	2·063	61°00′
10′	·4874	·8732	·5581	1·792	1·145	2·052	50′
20′	·4899	·8718	·5619	1·780	1·147	2·041	40′
30′	·4924	·8704	·5658	1·767	1·149	2·031	30′
40′	·4950	·8689	·5696	1·756	1·151	2·020	20′
50′	·4975	·8675	·5735	1·744	1·153	2·010	10,
30°00′	·5000	·8660	·5774	1·732	1·155	2·000	60°00′
10′	·5025	·8646	·5812	1·720	1·157	1·990	50′
20′	·5050	·8631	·5851	1·709	1·159	1·980	40′
30′	·5075	·8616	·5890	1·698	1·161	1·970	30′
40′	·5100	·8601	·5930	1·686	1·163	1·961	20′
50′	·5125	·8587	·5969	1·675	1·165	1·851	10′
31°00′	·5150	·8572	·6009	1·664	1·167	1·942	59°00′
10′	·5175	·8557	·6048	1·653	1·169	1·932	50′
20′	·5200	·8542	·6088	1·643	1·171	1·923	40′
30′	·5225	·8526	·6128	1·632	1·173	1·914	30′
40′	·5250	·8511	·6168	1·621	1·175	1·905	20′
50′	·5275	·8496	·6208	1·611	1·177	1·896	10′
Angle	cos	sin	cot	tan	csc	sec	Angle

Values of Trigonometric Functions *(Contd.)*

Angle	sin	cos	tan	cot	sec	csc	Angle
32°00′	·5299	·8480	·6249	1·600	1·179	1·887	58°00′
10′	·5324	·8465	·6289	1·590	1·181	1·878	50′
20′	·5348	·8450	·6330	1·580	1·184	1·870	40′
30′	·5373	·8434	·6371	1·570	1·186	1·861	30′
40′	·5398	·8418	·6412	1·560	1·188	1·853	20′
50′	·5422	·8403	·6453	1·550	1·190	1·844	10′
33°00′	·5446	·8387	·6494	1·540	1·192	1·836	57°00′
10′	·5471	·8371	·6536	1·530	1·195	1·828	50′
20′	·5495	·8355	·6577	1·520	1·197	1·820	40′
30′	·5519	·8339	·6619	1·511	1·199	1·812	30′
40′	·5544	·8323	·6661	1·501	1·202	1·804	20′
50′	·5568	·8307	·6703	1·492	1·204	1·796	10′
34°00′	·5592	·8290	·6745	1·483	1·206	1·788	56°00′
10′	·5616	·8274	·6787	1·473	1·209	1·781	50′
20′	·5640	·8258	·6830	1·464	1·211	1·773	40′
30′	·5664	·8241	·6873	1·455	1·213	1·766	30′
40′	·5688	·8225	·6916	1·446	1·216	1·758	20′
50′	·5712	·8208	·6959	1·437	1·218	1·751	10′
35°00′	·5736	·8192	·7002	1·428	1·221	1·743	55°00′
10′	·5760	·8175	·7046	1·419	1·223	1·736	50′
20′	·5783	·8158	·7089	1·411	1·226	1·729	40′
30′	·5807	·8141	·7133	1·402	1·228	1·722	30′
40′	·5831	·8124	·7177	1·393	1·231	1·715	20′
50′	·5854	·8107	·7221	1·385	1·233	1·708	10′
Angle	cos	sin	cot	tan	csc	sec	Angle

Values of Trigonometric Functions

(Contd.)

Angle	sin	cos	tan	cot	sec	csc	Angle
36°00′	·5878	·8090	·7265	1·376	1·236	1·701	54°00′
10′	·5901	·8073	·7310	1·368	1·239	1·695	50′
20′	·5925	·8056	·7355	1·360	1·241	1·688	40′
30′	·5948	·8039	·7400	1·351	1·244	1·681	30′
40′	·5972	·8021	·7445	1·343	1·247	1·675	20′
50′	·5995	·8004	·7490	1·335	1·249	1·668	10′
37°00′	·6018	·7986	·7536	1·327	1·252	1·662	53°00′
10′	·6041	·7969	·7581	1·319	1·255	1·655	50′
20′	·6065	·7951	·7627	1·311	1·258	1·649	40′
30′	·6088	·7934	·7673	1·303	1·260	1·643	30′
40′	·6111	·7916	·7720	1·295	1·263	1·636	20′
50′	·6134	·7898	·7766	1·288	1·266	1·630	10′
38°00′	·6157	·7880	·7813	1·280	1·269	1·624	52°00′
10′	·6180	·7862	·7860	1·272	1·272	1·618	50′
20′	·6202	·7844	·7907	1·265	1·275	1·612	40′
30′	·6225	·7826	·7954	1·257	1·278	1·606	30′
40′	·6248	·7808	·8002	1·250	1·281	1·601	20′
50′	·6271	·7790	·8050	1·242	1·284	1·595	10′
39°00′	·6293	·7771	·8098	1·235	1·287	1·589	51°00′
10′	·6316	·7753	·8146	1·228	1·290	1·583	50′
20′	·6338	·7735	·8195	1·220	1·293	1·578	40′
30′	·6361	·7716	·8243	1·213	1·296	1·572	30′
40′	·6383	·7698	·8292	1·206	1·299	1·567	20′
50′	·6406	·7679	·8342	1·199	1·302	1·561	10′
Angle	cos	sin	cot	tan	csc	sec	Angle

Values of Trigonometric Functions *(Contd.)*

Angle	sin	cos	tan	cot	sec	csc	Angle
40°00′	·6428	·7670	·8391	1·192	1·305	1·556	50°00′
10′	·6450	·7642	·8441	1·185	1·309	1·550	50′
20′	·6472	·7623	·8491	1·178	1·312	1·545	40′
30′	·6494	·7604	·8541	1·171	1·315	1·540	30′
40′	·6517	·7585	·8591	1·164	1·318	1·535	20′
50′	·6539	·7566	·8642	1·157	1·322	1·529	10′
41°00′	·6561	·7547	·8693	1·150	1·325	1·524	49°00′
10′	·6583	·7528	·8744	1·144	1·328	1·519	50′
20′	·6604	·7509	·8796	1·137	1·332	1·514	40′
30′	·6626	·7490	·8847	1·130	1·335	1·509	30′
40′	·6648	·7470	·8899	1·124	1·339	1·504	20′
50′	·6670	·7451	·8952	1·117	1·342	1·499	10′
42°00′	·6691	·7431	·9004	1·111	1·346	1·494	48°00′
10′	·6713	·7412	·9057	1·104	1·349	1·490	50′
20′	·6734	·7392	·9110	1·098	1·353	1·485	40′
30′	·6756	·7373	·9163	1·091	1·356	1·480	30′
40′	·6777	·7353	·9217	1·085	1·360	1·476	20′
50′	·6799	·7333	·9271	1·079	1·364	1·471	10′
43°00′	·6820	·7314	·9325	1·072	1·367	1·466	47°00′
10′	·6841	·7294	·9380	1·066	1·371	1·462	50′
20′	·6862	·7274	·9435	1·060	1·375	1·457	40′
30′	·6884	·7254	·9490	1·054	1·379	1·453	30′
40′	·6905	·7234	·9545	1·048	1·382	1·448	20′
50′	·6926	·7214	·9601	1·042	1·386	1·444	10′
Angle	cos	sin	cot	tan	csc	sec	Angle

Values of Trigonometric Functions

(*Contd.*)

Angle	sin	cos	tan	cot	sec	csc	Angle
44°00′	·6947	·7193	·9657	1·036	1·390	1·440	46°00′
10′	·6967	·7173	·9713	1·030	1·394	1·435	50′
20′	·6988	·7153	·9770	1·024	1·398	1·431	40′
30′	·7009	·7133	·9827	1·018	1·402	1·427	30′
40′	·7030	·7112	·9884	1·012	1·406	1·423	20′
50′	·7050	·7092	·9942	1·006	1·410	1·418	10′
45°00′	·7071	·7071	1·000	1·000	1·414	1·414	45°00′
Angle	cos	sin	cot	tan	csc	sec	Angle

TABLE 2
Common Logarithms

N	0	1	2	3	4	5	6	7	8	9
10	·0000	·0043	·0086	·0128	·0170	·0212	·0253	·0294	·0334	·0374
11	·0414	·0453	·0492	·0531	·0569	·0607	·0645	·0682	·0719	·0755
12	·0792	·0828	·0864	·0899	·0934	·0969	·1004	·1038	·1072	·1106
13	·1139	·1173	·1206	·1239	·1271	·1303	·1335	·1367	·1399	·1430
14	·1461	·1492	·1523	·1553	·1584	·1614	·1644	·1673	·1703	·1732
15	·1761	·1790	·1818	·1847	·1875	·1903	·1931	·1959	·1987	·2014
16	·2041	·2068	·2095	·2122	·2148	·2175	·2201	·2227	·2253	·2279
17	·2304	·2330	·2355	·2380	·2405	·2480	·2455	·2480	·2504	·2529
18	·2553	·2577	·2601	·2625	·2648	·2672	·2695	·2718	·2742	·2765
19	·2788	·2810	·2833	·2856	·2878	·2900	·2923	·2945	·2967	·2989
20	·3010	·3032	·3054	·3075	·3096	·3118	·3139	·3160	·3181	·3201
21	·3222	·3243	·3263	·3284	·3304	·3324	·3345	·3365	·3385	·3404
22	·3424	·3444	·3464	·3483	·3502	·3522	·3541	·3560	·3579	·3598
23	·3617	·3636	·3655	·3674	·3692	·3711	·3729	·3747	·3766	·3784
24	·3802	·3820	·3838	·3856	·3874	·3892	·3909	·3927	·3945	·3962
25	·3979	·3997	·0414	·4031	·4048	·4065	·4082	·4099	·4116	·4133
26	·4150	·4166	·4183	·4200	·4216	·4232	·4249	·4265	·4281	·4298
27	·4314	·4330	·4346	·4362	·4378	·4393	·4409	·4425	·4440	·4456
28	·4472	·4487	·4502	·4518	·4533	·4548	·4564	·4579	·4594	·4609
29	·4624	·4639	·4654	·4669	·4683	·4698	·4713	·4728	·4742	·4757
30	·4771	·4786	·4800	·4814	·4829	·4843	·4857	·4871	·4886	·4900
31	·4914	·4928	·4942	·4955	·4969	·4983	·4997	·5011	·5024	·5038
32	·5051	·5065	·5079	·5092	·5105	·5119	·5132	·5145	·5159	·5172
33	·5185	·5198	·5211	·5224	·5237	·5250	·5263	·5276	·5289	·5302
34	·5315	·5328	·5340	·5353	·5366	·5378	·5391	·5403	·5416	·5428

Common Logarithms *(Contd.)*

N	0	1	2	3	4	5	6	7	8	9
35	·5441	·5453	·5465	·5478	·5490	·5502	·5514	·5527	·5539	·5551
36	·5563	·5575	·5587	·5599	·5611	·5623	·5635	·5647	·5658	·5670
37	·5682	·5694	·5705	·5717	·5729	·5740	·5752	·5763	·5775	·5786
38	·5798	·5809	·5821	·5832	·5843	·5855	·5866	·5877	·5888	·5899
39	·5911	·5922	·5933	·5944	·5955	·5966	·5977	·5988	·5999	·6010
40	·6021	·6031	·6042	·6053	·6064	·6075	·6085	·6006	·6107	·6117
41	·6128	·6138	·6149	·6160	·6170	·6180	·6191	·6201	·6212	·6222
42	·6232	·6243	·6253	·6263	·6274	·6284	·6294	·6304	·6314	·6325
43	·6335	·6345	·6355	·6365	·6375	·6385	·6395	·6405	·6415	·6425
44	·6435	·6444	·6454	·6464	·6474	·6484	·6493	·6503	·6513	·6522
45	·6532	·6542	·6551	·6561	·6571	·6580	·6590	·6599	·6609	·6618
46	·6628	·6637	·6646	·6656	·6665	·6675	·6684	·6693	·6702	·6712
47	·6721	·6730	·6739	·6749	·6758	·6767	·6776	·6785	·6794	·6803
48	·6812	·6821	·6830	·6839	·6848	·6857	·6866	·6875	·6884	·6893
49	·6802	·6911	·6920	·6928	·6937	·6946	·6955	·6964	·6972	·6891
50	·6990	·6998	·7007	·7016	·7024	·7033	·7042	·7050	·7059	·7067
51	·7076	·7084	·7093	·7101	·7110	·7118	·7126	·7135	·7143	·7152
52	·7160	·7168	·7177	·7185	·7193	·7202	·7210	·7218	·7226	·7235
53	·7243	·7251	·7259	·7267	·7275	·7284	·7292	·7300	·7308	·7316
54	·7324	·7332	·7340	·7348	·7356	·7364	·7372	·7380	·7388	·7396
55	·7404	·7412	·7419	·7427	·7435	·7443	·7451	·7459	·7466	·7474
56	·7482	·7490	·7497	·7505	·7513	·7520	·7528	·7536	·7543	·7551
57	·7559	·7566	·7574	·7582	·7589	·7597	·7604	·7612	·7619	·7627
58	·7634	·7612	·7649	·7657	·7664	·7672	·7679	·7686	·7694	·7701
59	·7709	·7712	·7723	·7731	·7738	·7745	·7752	·7760	·7767	·7774

Common Logarithms

(Contd.)

N	0	1	2	3	4	5	6	7	8	9
60	·7782	·7789	·7796	·7803	·7810	·7818	·7825	·7832	·7839	·7846
61	·7853	·7860	·7868	·7875	·7882	·7889	·7896	·7903	·7910	·7917
62	·7924	·7931	·7938	·7945	·7952	·7959	·7966	·7973	·7980	·7987
63	·7993	·8000	·8007	·8014	·8021	·8028	·8035	·8041	·8048	·8055
64	·8062	·8069	·8075	·8082	·8089	·8096	·8102	·8109	·8106	·8122
65	·8129	·8136	·8142	·8149	·8156	·8162	·8169	·8176	·8182	·8189
66	·8195	·8202	·8209	·8215	·8222	·8228	·8235	·8241	·8248	·8254
67	·8261	·8267	·8274	·8280	·8287	·8293	·8299	·8306	·8312	·8319
68	·8325	·8331	·8338	·8344	·8351	·8357	·8363	·8370	·8376	·8332
69	·8388	·8395	·8401	·8407	·8414	·8420	·8426	·8432	·8439	·8444
70	·8451	·8457	·8463	·8470	·8476	·8482	·8488	·8494	·8500	·8505
71	·8513	·8519	·8525	·8531	·8537	·8543	·8549	·8555	·8561	·8567
72	·8573	·8579	·8585	·8591	·8597	·8603	·8609	·8615	·8621	·8627
73	·8633	·8639	·8645	·8651	·8657	·8663	·8669	·8675	·8681	·8686
74	·8692	·8698	·8704	·8710	·8716	·8722	·8727	·8733	·8739	·8745
75	·8751	·8756	·8762	·8768	·8774	·8779	·8785	·8791	·8797	·8802
76	·8808	·8814	·8820	·8825	·8831	·8837	·8842	·8848	·8854	·8859
77	·8865	·8871	·8876	·8882	·8887	·8893	·8899	·8904	·8910	·8915
78	·8921	·8927	·8932	·8938	·8943	·8949	·8954	·8960	·8956	·8971
79	·8976	·8982	·8987	·8993	·8998	·9004	·9009	·9015	·9020	·9025
80	·9031	·9036	9042	·9047	·9053	·9058	·9063	·9069	·9074	·9079
81	·9085	·9090	·9096	·9101	·9106	·9112	·9117	·9122	·9128	·9133
82	·9138	9143	·9149	·9154	·9159	·9165	·9170	·9175	·9180	·9186
83	·9191	·9196	·9201	·9206	·9212	·9217	·9222	·9227	9232	·9238
84	·9243	·9248	·9253	·9258	·9263	·9269	·9274	·9279	·9284	·9289
85	·9294	·9299	·9304	·9309	·9315	·9320	·9325	·9330	·9335	·9340

Common Logarithms *(Contd.)*

N	0	1	2	3	4	5	6	7	8	9
86	·9345	·9350	·9355	·9360	·9365	·9370	·9375	·9380	·9385	·9390
87	·9395	·9400	·9405	·9410	·9415	·9420	·9425	·9430	·9435	·9440
88	·9445	·9450	·9455	·9460	·9465	·9469	·9474	·9479	·9484	·9489
89	·9494	·9499	·9504	·9509	·9513	·9518	·9523	·9528	·9533	·9538
90	·9542	·9547	·9552	·9557	·9562	·9566	·9571	·9576	·9581	·9586
91	·9590	·9595	·9600	·9605	·9609	·9614	·9619	·9624	·9628	·9633
92	·9638	·9643	·9647	·9652	·9657	·9661	·9666	·9671	·9675	·9680
93	·9685	·9689	·9694	·9699	·9703	·9708	·9713	·9717	·9722	·9727
94	·9731	·9736	·9741	·9745	·9750	·9754	·9759	·9763	·9768	·9773
95	·9777	·9782	·9786	·9791	·9795	·9800	·9805	·9809	·9814	·9818
96	·9823	·9827	·9832	·9836	·9841	·9845	·9850	·9854	·9859	·9863
97	·9868	·9872	·9877	·9881	·9886	·9890	·9894	·9899	·9903	·9908
98	·9912	·9917	·9921	·9926	·9930	·9934	·9939	·9943	·9948	·9952
99	·9956	·9961	·9965	·9969	·9974	·9978	·9983	·9987	·9991	·9996

TABLE 3
Four-Place Logarithms of Values of Trigonometric Functions*

Angle	L sin	L tan	L cot	L cos	Angle
0°00′	—	—	—	10·0000	90°00′
10′	7·4637	7·4637	12·5363	10·0000	50′
20′	7·7648	7·7648	12·2352	10·0000	40′
30′	7·9408	7·9409	12·0591	10·0000	30′
40′	8·0658	8·0658	11·9342	10·0000	20′
50′	8·1627	8·1627	11·8373	10·0000	10′
1°00′	8·2419	8·2419	11·7581	9·9999	89°00′
10′	8·3088	8·3089	11·6911	9·9999	50′
20′	8·3668	8·3669	11·6331	9·9999	40′
30′	8·4179	8·4181	11·5819	9·9999	30′
40′	8·4637	8·4638	11·5362	9·9998	20′
50′	8·5050	8·5053	11·4947	9·9998	10′
2°00′	8·5428	8·5431	11·4569	9·9997	88°00′
10′	8·5776	8·5779	11·4221	9·9997	50′
20′	8·6097	8·6101	11·3899	9·9996	40′
30′	8·6397	8·6401	11·3599	9·9996	30′
40′	8·6677	8·6682	11·3318	9·9995	20′
50′	8·6940	8·6945	11·3055	9·9995	10′
3°00′	8·7188	8·7194	11·2806	9·9994	87°00′
10′	8·7423	8·7429	11·2571	9·9993	50′
20′	8·7645	8·7652	11·2348	9·9993	40′
30′	8·7857	8·7865	11·2135	9·9992	30′
40′	8·8059	8·8067	11·1933	9·9991	20′
50′	8·8251	8·8261	11·1739	9·9990	10′
Angle	L cos	L cot	L tan	L sin	Angle

*This table gives the logarithms increased by 10. Hence in each case 10 should be subtracted.

Four-Place Logarithms of Values of Trigonometric Functions *(Contd.)*

Angle	L sin	L tan	L cot	L cos	Angle
4°00′	8·8436	8·8446	11·1554	9·9989	86°00′
10′	8·8613	8·8624	11·1376	9·9989	50′
20′	8·8783	8·8795	11·1205	9·9988	40′
30′	8·8946	8·8960	11·1040	9·9987	30′
40′	8·9104	8·9118	11·0882	9·9986	20′
50′	8·9256	8·9272	11·0728	9·9985	10′
5°00′	8·9403	8·9420	11·0580	9·9983	85°00′
10′	8·9545	8·9563	11·0437	9·9982	50′
20′	8·9682	8·9701	11·0299	9·9981	40′
30′	8·9816	8·9836	11·0164	9·9980	30′
40′	8·9945	8·9966	11·0034	9·9979	20′
50′	9·0070	9·0093	10·9907	9·9977	10′
6°00′	9·0192	9·0216	10·9784	9·9976	84°00′
10′	9·0311	9·0336	10·9664	9·9975	50′
20′	9·0426	9·0453	10·9547	9·9973	40′
30′	9·0539	9·0567	10·9433	9·9972	30′
40′	9·0648	9·0678	10·9322	9·9971	20′
50′	9·0755	9·0786	10·9214	9·9969	10′
7°00′	9·0859	9·0891	10·9109	9·9968	83°00′
10′	9·0961	9·0995	10·9005	9·9966	50′
20′	9·1060	9·1096	10·8904	9·9964	40′
30′	9·1157	9·1194	10·8806	9·9963	30′
40′	9·1252	9·1291	10·8709	9·9961	20′
50′	9·1345	9·1385	10·8615	9·9959	10′
Angle	L cos	L cot	L tan	L sin	Angle

Four-Place Logarithms of Values of Trigonometric Functions (*Contd.*)

Angle	L sin	L tan	L cot	L cos	Angle
8°00′	9·1436	9·1478	10·8522	9·9958	82°00′
10′	9·1525	9·1569	10·8431	9·9956	50′
20′	9·1612	9·1658	10·8342	9·9954	40′
30′	9·1697	9·1745	10·8255	9·9952	30′
40′	9·1781	9·1831	10·8169	9·9950	20′
50′	9·1863	9·1915	10·8085	9·9948	10′
9°00′	9·1943	9·1997	10·8003	9·9946	81°00′
10′	9·2022	9·2078	10·7922	9·9944	50′
20′	9·2100	9·2158	10·7842	9·9942	40′
30′	9·2176	9·2236	10·7764	9·9940	30′
40′	9·2251	9·2313	10·7687	9·9938	20′
50′	9·2324	9·2389	10·7611	9·9936	10′
10°00′	9·2397	9·2463	10·7537	9·9934	80°00′
10′	9·2468	9·2536	10·7464	9·9931	50′
20′	9·2538	9·2609	10·7391	9·9929	40′
30′	9·2606	9·2680	10·7320	9·9927	30′
40′	9·2674	9·2750	10·7250	9·9924	20′
50′	9·2740	9·2819	10·7181	9·9922	10′
11°00′	9·2806	9·2887	10·7113	9·9919	79°00′
10′	9·2870	9·2953	10·7047	9·9917	50′
20′	9·2934	9·3020	10·6980	9·9914	40′
30′	9·2997	9·3085	10·6915	9·9912	30′
40′	9·3058	9·3149	10·6851	9·9909	20′
50′	9·3119	9·3112	10·6788	9·9907	10′
Angle	L cos	L cot	L tan	L sin	Angle

Four-Place Logarithms of Values of Trigonometric Functions *(Contd.)*

Angle	L sin	L tan	L cot	L cos	Angle
12°00′	9·3179	9·3275	10·6725	9·9904	78°00′
10′	9·3238	9·3336	10·6664	9·9901	50′
20′	9·3296	9·3397	10·6603	9·9899	40′
30′	9·3353	9·3458	10·6542	9·9896	30′
40′	9·3410	9·3517	10·6483	9·9893	20′
50′	9·3466	9·3576	10·6424	9·9890	10′
13°00′	9·3521	9·3634	10·6366	9·9887	77°00′
10′	9·3575	9·3691	10·6309	9·9884	50′
20′	9·3629	9·3748	10·6252	9·9881	40′
30′	9·3682	9·3804	10·6196	9·9878	30′
40′	9·3734	9·3859	10·6141	9·9875	20′
50′	9·3786	9·3914	10·6086	9·9872	10′
14°00′	9·3837	9·3968	10·6032	9·9869	76°00′
10′	9·3887	9·4021	10·5979	9·9866	50′
20′	9·3937	9·4074	10·5926	9·9863	40′
30′	9·3986	9·4127	10·5873	9·9859	30′
40′	9·4035	9·4178	10·5822	9·9856	20′
50′	9·4083	9·4230	10·5770	9·9853	10′
15°00′	9·4130	9·4281	10·5719	9·9849	75°00′
10′	9·4177	9·4331	10·5669	9·9846	50′
20′	9·4223	9·4381	10·5619	9·9843	40′
30′	9·4269	9·4430	10·5570	9·9839	30′
40′	9·4314	9·4479	10·5521	9·9836	20′
50′	9·4359	9·4527	10·5473	9·9832	10′
Angle	L cos	L cot	L tan	L sin	Angle

Four-Place Logarithms of Values of Trigonometric Functions *(Contd.)*

Angle	L sin	L tan	L cot	L cos	Angle
16°00′	9·4403	9·4575	10·5425	9·9828	74°00′
10′	9·4447	9·4622	10·5378	9·9825	50′
20′	9·4491	9·4669	10·5331	9·9821	40′
30′	9·4533	9·4716	10·5284	9·9817	30′
40′	9·4576	9·4762	10·5238	9·9814	20′
50′	9·4618	9·4808	10·5192	9·9810	10′
17°00′	9·4659	9·4853	10·5147	9·9806	73°00′
10′	9·4700	9·4898	10·5102	9·9802	50′
20′	9·4741	9·4943	10·5057	9·9798	40′
30′	9·4781	9·4987	10·5013	9·9794	30′
40′	9·4821	9·5031	10·4969	9·9790	20′
50′	9·4861	9·5075	10·4925	9·9786	10′
18°00′	9·4900	9·5118	10·4882	9·9782	72°00′
10′	9·4939	9·5161	10·4839	9·9778	50′
20′	9·4977	9·5203	10·4797	9·9774	40′
30′	9·5015	9·5245	10·4755	9·9770	30′
40′	9·5052	9·5287	10·4713	9·9765	20′
50′	9·5090	9·5329	10·4671	9·9761	10′
19°00′	9·5126	9·5370	10·4630	9·9757	71°00′
10′	9·5163	9·5411	10·4589	9·9752	50′
20′	9·5199	9·5451	10·4549	9·9748	40′
30′	9·5235	9·5491	10·4509	9·9743	30′
40′	9·5270	9·5531	10·4469	9·9739	20′
50′	9·5306	9·5571	10·4429	9·9734	10′
Angle	L cos	L cot	L tan	L sin	Angle

Four-Place Logarithms of Values of Trigonometric Functions (*Contd.*)

Angle	L sin	L tan	L cot	L cos	Angle
20°00′	9·5341	9·5611	10·4389	9·9730	70°00′
10′	9·5375	9·5650	10·4350	9·9725	50′
20′	9·5409	9·5689	10·4311	9·9721	40′
30′	9·5443	9·5727	10·4273	9·9716	30′
40′	9·5477	9·5766	10·4234	9·9711	20′
50′	9·5510	9·5804	10·4196	9·9706	10′
21°00′	9·5543	9·5842	10·4158	9·9702	69°00′
10′	9·5576	9·5879	10·4121	9·9697	50′
20′	9·5609	9·5917	10·4083	9·9692	40′
30′	9·5641	9·5954	10·4046	9·9687	30′
40′	9·5673	9·5991	10·4009	9·9682	20′
50′	9·5704	9·6028	10·3972	9·9677	10′
22°00′	9·5736	9·6064	10·3936	9·9672	68°00′
10′	9·5767	9·6100	10·3900	9·9667	50′
20′	9·5798	9·6136	10·3864	9·9661	40′
30′	9·5828	9·6172	10·3828	9·9656	30′
40′	9·5859	9·6208	10·3792	9·9651	20′
50′	9·5889	9·6243	10·3757	9·9646	10′
23°00′	9·5919	9·6279	10·3721	9·9640	67°00′
10′	9·5948	9·6314	10·3686	9·9635	50′
20′	9·5978	9·6348	10·3652	9·9629	40′
30′	9·6007	9·6383	10·3617	9·9624	30′
40′	9·6036	9·6417	10·3583	9·9518	20′
50′	9·6065	9·6452	10·3548	9·9613	10′
Angle	L cos	L cot	L tan	L sin	Angle

Four-Place Logarithms of Values of Trigonometric Functions (*Contd.*)

Angle	L sin	L tan	L cot	L cos	Angle
24°00′	9·6093	9·6486	10·3514	9·9607	66°00′
10′	9·6121	9·6520	10·3480	9·9602	50′
20′	9·6149	9·6553	10·3447	9·9596	40′
30′	9·6177	9·6587	10·3413	9·9590	30′
40′	9·6205	9·6620	10·3380	9 9584	20′
50′	9·6232	9·6654	10 3346	9·9579	10′
25°00′	9·6259	9·6687	10·3313	9·9573	65°00′
10′	9·6286	9·6720	10·3280	9·9567	50′
20′	9·6313	9·6752	10·3248	9·9561	40′
30′	9·6340	9·6785	10·3215	9·9555	30′
40′	9·6366	9·6817	10·3183	9·9549	20′
50′	9·6392	9·6850	10·3150	9·9543	10′
26°00′	9·6418	9·6882	10·3118	9·9537	64°00′
10′	9·6444	9·6914	10·3086	9·9530	50′
20′	9·6470	9·6946	10·3054	9·9524	40′
30′	9·6495	9·6977	10·3023	9·9518	30′
40′	9·6521	9·7009	10·2991	9·9512	20′
50′	9·6546	9·7040	10·2960	9·9505	10′
27°00′	9·6570	9·7072	10·2928	9·9499	63°00′
10′	9·6595	9·7103	10·2897	9·9492	50′
20′	9·6620	9·7134	10·2866	9·9486	40′
30′	9·6644	9·7165	10·2835	9·9479	30′
40′	9·6668	9·7196	10·2804	9·9473	20′
50′	9·6692	9·7226	10·2774	9·9466	10′
Angle	L cos	L cot	L tan	L sin	Angle

Four-Place Logarithms of Values of Trigonometric Functions (*Contd.*)

Angle	L sin	L tan	L cot	L cos	Angle
28°00′	9·6716	9·7257	10·2743	9·9459	62°00′
10′	6·6740	9·7287	10·2713	9·9453	50′
20′	9·6763	9·7317	10·2683	9·9446	40′
30′	9·6787	9·7348	10·2652	9·9439	30′
40′	9·6810	9·7378	10·2622	9·9432	20′
50′	9·6833	9·7408	10·2592	9·9425	10′
29°00′	9·6856	9·7438	10·2562	9·9418	61°00′
10′	9·6878	9·7467	10·2533	9·9411	50′
20′	9·6901	9·7497	10·2503	9·9404	40′
30′	9·6923	9·7526	10·2474	9·9397	30′
40′	9·6946	9·7556	10·2444	9·9390	20′
50′	9·6968	9·7585	17·2415	9·9383	10′
30°00′	9·6990	9·7614	10·2386	9·9375	60°00′
10′	9·7012	9·7644	10·2356	9·9368	50′
20′	9·7033	9·7673	10·2327	9·9361	40′
30′	9·7055	9·7701	10·2299	9·9353	30′
40′	9·7076	9·7730	10·2270	9·9346	20′
50′	9·7097	9·7759	10·2241	9·9338	10′
31°00′	9·7118	9·7788	10·2212	9·9331	59°00′
10′	9·7139	9·7816	10·2184	9·9323	50′
20′	9·7160	9·7845	10·2155	9·9315	40′
30′	9·7181	9·7873	10·2127	9·9308	30′
40′	9·7201	9·7902	10·2098	9·9300	20′
50′	9·7222	9·7930	10·2070	9·9292	10′
Angle	L cos	L cot	L tan	L sin	Angle

Four-Place Logarithms of Values of Trigonometric Functions *(Contd.)*

Angle	L sin	L tan	L cot	L cos	Angle
32°00′	9·7242	9·7958	10·2042	9·9284	58°00′
10′	9·7262	9·7986	10·2014	9·9276	50′
20′	9·7282	9·8014	10·1986	9·9268	40′
30′	9·7302	9·8042	10·1958	9·9260	30′
40′	9·7322	9·8070	10·1930	9·9252	20′
50′	9·7342	9·8097	10·1903	9·9244	10′
33°00′	9·7361	9·8125	10·1875	9·9236	57°00′
10′	9·7380	9·8153	10·1847	9·9228	50′
20′	9·7400	9·8180	10·1820	9·9219	40′
30′	9·7419	9·8208	10·1792	9·9211	30′
40′	9·7438	9·8235	10·1765	9·9203	20′
50′	9·7457	9·8263	10·1737	9·9194	10′
34°00′	9·7476	9·8290	10·1710	9·9186	56°00′
10′	9·7494	9·8317	10·1683	9·9177	50′
20′	9·7513	9·8344	10·1656	9·9169	40′
30′	9·7531	9·8371	10·1629	9·9160	30′
40′	9·7550	9·8398	10·1602	9·9151	20′
50′	9·7568	9·8425	10·1575	9·9142	10′
35°00′	9·7586	9·8452	10·1548	9·9134	55°00′
10′	9·7604	9·8479	10·1521	9·9125	50′
20′	9·7622	9·8506	10·1494	9·9116	40′
30′	9·7640	9·8533	10·1467	9·9107	30′
40′	9·7657	9·8559	10·1441	9·9098	20′
50′	9·7675	9·8586	10·1414	9·9089	10′
Angle	L cos	L cot	L tan	L sin	Angle

Four-Place Logarithms of Values of Trigonometric Functions *(Contd.)*

Angle	L sin	L tan	L cot	L cos	Angle
36°00′	9·7692	9·8613	10·1387	9·9080	54°00′
10′	9·7710	9·8639	10·1361	9·9070	50′
20′	9·7727	9·8666	10·1334	9·9061	40′
30′	9·7744	9·8692	10·1308	9·9052	30′
40′	9·7761	9·8718	10·1282	9·9042	20′
50′	9·7778	0·8745	10·1255	9·9033	10′
37°00′	9·7795	9·8771	10·1229	9·9023	53°00′
10′	9·7811	9·8797	10·1203	9·9014	50′
20′	9·7828	9·8824	10·1176	9·9004	40′
30′	9·7844	9·8850	10·1150	9·8995	30′
40′	9·7861	9·8876	10·1124	9·8985	20′
50′	9·7877	9·8902	10·1098	9·8975	10′
38°00′	9·7893	9·8928	10·1072	9·8965	52°00′
10′	9·7910	9·8945	10·1046	9·8955	50′
20′	9·7926	9·8980	10·1020	9·8945	40′
30′	9·7941	9·9006	10·0994	9·8935	30′
40′	9·7957	9·9032	10·0968	9·8925	20′
50′	9·7973	9·9058	10·0942	9·8915	10′
39°00′	9·7989	9·9084	10·0916	9·8905	51°00′
10′	9·8004	9·9110	10·0890	9·8895	50′
20′	9·8020	9·9135	10·0865	9·8884	40′
30′	9·8035	9·9161	10·0839	9·8874	30′
40′	9·8053	9·9187	10·0813	9·8864	20′
50′	9·8066	9·9212	10·0788	9·8853	10′
Angle	L cos	L cot	L tan	L sin	Angle

Four-Place Logarithms of Values of Trigonometric Functions (*Contd.*)

Angle	L sin	L tan	L cot	L cos	Angle
40°00′	9·8081	9·9238	10·0762	9·8843	50°00′
10′	9·8096	9·9264	10·0736	9·8832	50′
20′	9·8111	9·9289	10.0711	9·8821	40′
30′	9·8125	9·9315	10·0685	9·8810	30′
40′	9·8140	9·9341	10·0659	9·8800	20′
50′	9·8155	9·9366	10·0634	9·8789	10′
41°00′	9·8169	9·9392	10·0608	9·8778	49°00′
10′	9·8184	9·9417	10·0583	9·8767	50′
20′	9·8198	9·9443	10·0557	9·8756	40′
30′	9·8213	9·9468	10·0532	9·8745	30′
40′	9·8227	9·9494	10·0506	9·8733	20′
50′	9·8241	9·9519	10·0481	9·8722	10′
42°00′	9·8255	9·9544	10·0456	9·8711	48°00′
10′	9·8269	9·9570	10·0430	9·8699	50′
20′	9·8283	9·9595	10·0405	9·8688	40′
30′	9·8297	9·9621	10·0379	9·8676	30′
40′	9·8311	9·9646	10·0354	9·8665	20′
50′	9·8324	9·9671	10·0329	9·8653	10′
43°00′	9·8338	9·9697	10·0303	9·8641	47°00′
10′	9·8351	9·9722	10·0278	9·8629	50′
20′	9·8365	9·9747	10·0253	9·8618	40′
30′	9·8378	9·9772	10·0228	9·8606	30′
40′	9·8391	9·9798	10·0202	9·8594	20′
50′	9·8405	9·9823	10·0177	9·8582	10′
Angle	**L cos**	**L cot**	**L tan**	**L sin**	**Angle**

Four-Place Logarithms of Values of Trigonometric Functions (*Contd.*)

Angle	L sin	L tan	L cot	L cos	Angle
44°00′	9·8418	9·8848	10·0152	9·8569	46°00′
10′	9·8431	9·9874	10·0126	9·8557	50′
20′	9·8444	9·9899	10·0101	9·8545	40′
30′	9·8457	9·9924	10·0076	9·8532	30′
40′	9·8469	9·9949	10·0051	9·8520	20′
50′	9·8482	9·9975	10·0025	9·8507	10′
45°00′	9·8495	10·0000	10·0000	9·8495	45°00′
Angle	L cos	L cot	L tan	L sin	Angle

TABLE 4
Squares and Square Roots

N	N²	√N	√10N	N	N²	√N	√10N
1.0	1.00	1.000	3.162	3.5	12.25	1.871	5.916
1.1	1.21	1.049	3.317	3.6	12.96	1.897	6.000
1.2	1.44	1.095	3.464	3.7	13.69	1.924	6.083
1.3	1.69	1.140	3.606	3.8	14.44	1.949	6.164
1.4	1.96	1.183	3.742	3.9	15.21	1.975	6.245
1.5	2.25	1.225	3.873	4.0	16.00	2.000	6.325
1.6	2.56	1.265	4.000	4.1	16.81	2.025	6.403
1.7	2.89	1.304	4.123	4.2	17.64	2.049	6.481
1.8	3.24	1.342	4.243	4.3	18.49	2.074	6.557
1.9	3.61	1.378	4.359	4.4	18.36	2.098	6.633
2.0	4.00	1.414	4.472	4.5	20.25	2.121	6.708
2.1	4.41	1.449	4.583	4.6	21.16	2.145	6.782
2.2	4.84	1.483	4.690	4.7	22.09	2.168	6.856
2.3	5.29	1.517	4.796	4.8	23.04	2.191	6.928
2.4	5.76	1.549	4.899	4.9	24.01	2.214	7.000
2.5	6.25	1.581	5.000	5.0	25.00	2.236	7.071
2.6	6.76	1.612	5.099	5.1	26.01	2.258	7.141
2.7	7.29	1.643	5.196	5.2	27.04	2.280	7.211
2.8	7.84	1.673	5.292	5.3	28.09	2.302	7.280
2.9	8.41	1.703	5.385	5.4	29.16	2.324	7.348
3.0	9.00	1.732	5.477	5.5	30.25	2.345	7.416
3.1	9.61	1.761	5.568	5.6	31.36	2.366	7.483
3.2	10.24	1.789	5.657	5.7	32.49	2.387	7.550
3.3	10.89	1.817	5.745	5.8	33.64	2.408	7.616
3.4	11.56	1.844	5.831	5.9	34.81	2.429	7.681

Squares and Square Roots (*Contd.*)

N	N²	√N	√10N	N	N²	√N	√10N
6·0	36·00	2·449	7·746	8·0	64·00	2·828	8·944
6·1	37·21	2·470	7·810	8·1	65·61	2·846	9·000
6·2	38·44	2·490	7·870	8·2	67·24	2·864	9·055
6·3	39·69	2·510	7·937	8·3	68·89	2·881	9·110
6·4	40·96	2·530	8·000	8·4	70·56	2·898	9·165
6·5	42·25	2·550	8·062	8·5	72·25	2·915	9·220
6·6	43·56	2·569	8·124	8·6	73·96	2·933	9·274
6·7	44·89	2·583	8·185	8·7	75·69	2·950	9·327
6·8	46·24	2·608	8·246	8·8	77·44	2·966	9·381
6·9	47·61	2·627	8·307	8·9	79·21	2·983	9·434
7·0	49·00	2·646	8·367	9·0	81·00	3·000	9·487
7·1	50·41	2·665	8·426	9·1	82·81	3·017	9·539
7·2	51·84	2·683	8·485	9·2	84·64	3·033	9·592
7·3	53·29	2·702	8·544	9·3	86·49	3·050	9·644
7·4	54·76	2·720	8·602	9·4	88·36	3·066	9·695
7·5	56·25	2·739	8·660	9·5	90·25	3·082	9·747
7·6	57·76	2·757	8·718	9·5	92·16	3·098	9·798
7·7	59·29	2·775	8·775	9·7	94·09	3·114	9·849
7·8	60·84	2·793	8·832	9·8	96·04	3·130	9·899
7·9	62·41	2·811	8·888	9·9	98·01	3·146	9·950
				10·0	10·000	3·162	1·000

TABTE 5
Cubes and Cube Roots

N	N^3	$\sqrt[3]{N}$	$\sqrt[3]{10N}$	$\sqrt[3]{100N}$	N	N^3	$\sqrt[3]{N}$	$\sqrt[3]{10N}$	$\sqrt[3]{100N}$
1·0	1·000	1·000	2·154	4·642	3·5	42·875	1·518	3·271	7·047
1·1	1·331	1·032	2·224	4·791	3·6	46·656	1·533	3·302	7·114
1·2	1·728	1·063	2·289	4·932	3·7	50·653	1·547	3·332	7·179
1·3	2·197	1·091	2·351	5·066	3·8	54·872	1·560	3·362	7·243
1·4	2·744	1·119	2·410	5·192	3·9	59·319	1·574	3·391	7·306
1·5	3·375	1·145	2·466	5·313	4·0	64·000	1·587	3·420	7·368
1·6	4·096	1·170	2·520	5·429	4·1	68·921	1·601	3·448	7·429
1·7	4·913	1·193	2·571	5·540	4·2	74·088	1·613	3·476	7·488
1·8	5·832	1·216	2·621	5·646	4·3	79·507	1·626	3·503	7·548
1·9	6·859	1·239	2·668	5·749	4·4	85·184	1·639	3·530	7·606
2·0	8·000	1·260	2·714	5·848	4·5	91·125	1·651	3·557	7·663
2·1	9·261	1·281	2·759	5·944	4·6	97·336	1·663	3·583	7·719
2·2	10·648	1·301	2·802	6·037	4·7	103·823	1·675	3·609	7·775
2·3	12·167	1·320	2·844	6·127	4·8	110·592	1·687	3·634	7·830
2·4	13·824	1·339	2·884	6·214	4·9	117·649	1·698	3·659	7·884
2·5	15·625	1·357	2·924	6·300	5·0	125·000	1·710	3·684	7·937
2·6	17·576	1·375	2·962	6·383	5·1	132·651	1·721	3·708	7·990
2·7	19·683	1·392	3·000	6·463	5·2	140·608	1·732	3·733	8·041
2·8	21·952	1·409	3·037	6·542	5·3	148·877	1·744	3·756	8·093
2·9	24·389	1·426	3·072	6·619	5·4	157·464	1·754	3·780	8·143
3·0	27·000	1·442	3·107	6·694	5·5	166·375	1·765	3·803	8·193
3·1	29·791	1·458	3·141	6·768	5·6	175·616	1·776	3·826	8·243
3·2	32·768	1·474	3·175	6·840	5·7	185·193	1·786	3·849	8·291
3·3	35·937	1·489	3·208	6·910	5·8	195·112	1·797	3·871	8·340
3·4	39·304	1·504	3·240	6·980	5·9	205·379	1·807	3·893	8·387

Cubes and Cube Roots

N	N³	$\sqrt[3]{N}$	$\sqrt[3]{10N}$	$\sqrt[3]{100N}$	N	N³	$\sqrt[3]{N}$	$\sqrt[3]{10N}$	$\sqrt[3]{100N}$
6.0	216.000	1.817	3.915	8.434	8.0	512.000	2.000	4.309	9.283
6.1	226.981	1.827	3.936	8.481	8.1	531.441	2.008	4.327	9.322
6.2	238.328	1.837	8.958	8.527	8.2	551.368	2.017	4.344	9.360
6.3	250.047	1.847	3.979	8.573	8.3	571.787	2.025	4.362	9.398
6.4	262.144	1.857	4.000	8.618	8.4	592.704	2.033	4.380	9.435
6.5	274.625	1.866	4.021	8.662	8.5	614.125	2.041	4.357	9.473
6.6	287.496	1.876	4.041	8.707	8.6	636.056	2.049	4.414	9.510
6.7	300.763	1.885	4.062	8.750	8.7	658.503	2.057	4.431	9.546
6.8	314.432	1.895	4.082	8.794	8.8	681.472	2.065	4.448	9.583
6.9	328.509	1.904	4.102	8.837	8.9	704.969	2.072	4.465	9.619
7.0	343.000	1.913	4.121	8.879	9.0	729.000	2.080	4.481	9.655
7.1	357.911	1.922	4.141	8.921	9.1	753.571	2.088	4.498	9.691
7.2	373.248	1.931	4.160	8.963	9.2	778.688	2.095	4.514	9.726
7.3	309.017	1.940	4.179	9.004	9.3	804.357	2.103	4.531	9.761
7.4	405.224	1.949	4.198	9.045	9.4	830.584	2.110	4.547	9.796
7.5	421.875	1.957	4.217	9.086	9.5	857.375	2.118	4.563	9.830
7.6	438.976	1.966	4.236	9.126	9.6	884.736	2.125	4.579	9.865
7.7	456.533	1.975	4.254	9.166	9.7	912.673	2.133	4.595	9.899
7.8	474.552	1.983	4.273	9.205	9.8	941.192	2.140	4.610	9.933
7.9	493.039	1.992	4.291	9.244	9.9	970.299	2.147	4.626	9.967
					10.0	1000.000	2.154	4.642	0.000

APPENDIX I

Binomial Probabilities

Entries in the table are the values of $b(x\,;\,n, p)$ for $n=1\cdots2, 10$; for $x : 0 \leqslant x \leqslant n$, for $p=$ ·01, ·05, ·10, ·15, ·20, ·25, ·30, $\frac{1}{3}$, ·35, ·40, ·45, ·49 and ·50.

n	x	·01	·05	·10	·15	·20	·25	·30	$\frac{1}{3}$	·35	·40	·45	·49	·50
2	0	·9801	·9025	·8100	·7225	·6400	·5625	·4900	·4444	·4225	·3600	·3025	·2601	·2500
	1	·0198	·0950	·1800	·2550	·3200	·3750	·4200	·4444	·4550	·4800	·4950	·4998	·5000
	2	·0001	·0025	·0100	·0225	·0400	·0625	·0900	·1111	·1225	·1600	·2025	·2401	·2500
3	0	·9703	·8574	·7290	·6141	·5120	·4219	·3430	·2963	·2746	·2160	·1664	·1327	·1250
	1	·0294	·1354	·2430	·3251	·3840	·4219	·4410	·4444	·4436	·4320	·4084	·3823	·3750
	2	·0003	·0071	·0270	·0574	·0960	·1406	·1890	·2222	·2389	·2880	·3341	·3674	·3750
	3	·0000	·0001	·0010	·0034	·0080	·0156	·0270	·0370	·0429	·0640	·0911	·1176	·1250
4	0	·9606	·8145	·6561	·5220	·4096	·3164	·2401	·1975	·1785	·1296	·0915	·0677	·0625
	1	·0388	·1715	·2916	·3685	·4096	·4219	·4116	·3951	·3845	·3456	·2995	·2600	·2500
	2	·0006	·0135	·0486	·0975	·1536	·2109	·2646	·2963	·3105	·3456	·3675	·3747	·3750
	3	·0000	·0005	·0036	·0115	·0256	·0469	·0756	·0988	·1115	·1536	·2005	·2400	·2500
	4	·0000	·0000	·0001	·0005	·0016	·0039	·0081	·0123	·0150	·0256	·0410	·0576	·0625
5	0	·9510	·7738	·5905	·4437	·3277	·2373	·1681	·1317	·1160	·0778	·0503	·0345	·0312
	1	·0480	·2036	·3280	·3915	·4096	·3955	·3602	·3292	·3124	·2592	·2059	·1657	·1562
	2	·0010	·2214	·0729	·1382	·2048	·2637	·3087	·3292	·3364	·3456	·3369	·3185	·3125
	3	·0000	·0011	·0081	·0244	·0512	·0879	·1323	·1646	·1811	·2304	·2757	·3060	·3125
	4	·0000	·0000	·0004	·0022	·0064	·0146	·0284	·0412	·0483	·0768	·1128	·1470	·1562
	5	·0000	·0000	·0000	·0001	·0003	·0010	·0024	·0041	·0153	·0102	·0185	·0283	·0312

Binomial Probabilities (Contined)

6	0	·9415	·7351	·5314	·3771	·2621	·1780	·1176	·0878	·0754	·0467	·0277	·0176	·0156
	1	·0571	·2321	·3543	·3993	·3932	·3560	·3025	·2634	·2437	·1866	·1359	·1014	·0938
	2	·0014	·0305	·0984	·1762	·2458	·2966	·3241	·3292	·3280	·3110	·2780	·2437	·2344
	3	·0000	·0021	·0146	·0415	·0819	·1318	·1852	·2195	·2355	·2765	·3032	·3121	·3125
	4	·0000	·0001	·0012	·0055	·0154	·0330	·0595	·0823	·0951	·1382	·1861	·2249	·2344
	5	·0000	·0000	·0001	·0004	·0015	·0044	·0102	·0165	·0205	·0369	·0609	·0864	·0938
	6	·0000	·0000	·0000	·0000	·0001	·0002	·C007	·0014	·0018	·0041	·0083	·0139	·0150
7	0	·9321	·6983	·4783	·3206	·2097	·1335	·0824	·0585	·0490	·0280	·0152	·0090	·0078
	1	·0659	·2573	·3720	·3960	·3610	·3115	·2471	·2048	·1848	·1306	·0872	·0603	·0547
	2	·0020	·0406	·1240	·2097	·2753	·3115	·3177	·3073	·2985	·2613	·2140	·1740	·1641
	3	·0000	·0036	·0230	·0617	·1147	·1730	·2269	·2561	·2679	·2903	·2918	·2786	·2734
	4	·0000	·0002	·0026	·0109	·0287	·0577	·0972	·1280	·1442	·1935	·2388	·2676	·2734
	5	·0000	·0000	·0002	·0012	·0043	·0115	·0250	·0384	·0466	·0774	·1172	·1543	·1641
	6	·0000	·0000	·0000	·0001	·0004	·0013	·0036	·0064	·0084	·0172	·0320	·0494	·0547
	7	·0000	·0000	·0000	·0000	·0000	·0001	·0002	·0005	·0006	·0016	·0037	·0068	·0078
8	0	·9227	·6634	·4305	·2725	·1678	·1001	·0576	·0390	·0319	·0168	·0084	·0046	·0039
	1	·0746	·2793	·3826	·3847	·3355	·2670	·1977	·1561	·1373	·0896	·0548	·0352	·0312
	2	·0026	·0515	·1488	·2376	·2936	·3115	·2965	·2731	·2587	·2090	·1569	·1183	·1094
	3	·0001	·0054	·0331	·0839	·1468	·2076	·2541	·2731	·2786	·2787	·2568	·2273	·2188
	4	·0000	·0004	·0046	·0185	·0459	·0865	·1361	·1707	·1876	·2322	·2627	·2730	·2734
	5	·0000	·0000	·0004	·0026	·0092	·0231	·0467	·0683	·0808	·1239	·1719	·2098	·2188
	6	·0000	·0000	·C000	·0002	·0011	·0038	·0100	·0171	·0217	·0413	·0703	·1008	·1094
	7	·0000	·0000	·0000	·0000	·0001	·0004	·0012	·0024	·0033	·0079	·0164	·0277	·0312
	8	·0000	·0000	·0000	·0000	·0000	·0000	·0001	·0002	·0002	·0007	·0017	·0033	·0039

Binomial Probabilities (*Continued*)

9	0	·9135	·6302	·3874	·2316	·1342	·0751	·0404	·0260	·0207	·0101	·0046	·0023	·0020
	1	·0830	·2985	·3874	·3679	·3020	·2253	·1556	·1171	·1004	·0605	·0339	·0202	·0176
	2	·0034	·0629	·1722	·2597	·3020	·3003	·2668	·2341	·2162	·1612	·1110	·0776	·0703
	3	·0001	·0077	·0446	·1069	·1762	·2336	·2668	·2731	·2716	·2508	·2119	·1739	·1641
	4	·0000	·0006	·0074	·0283	·0661	·1168	·1715	·2048	·2194	·2508	·2600	·2506	·2461
	5	·0000	·0000	·0008	·0050	·0165	·0389	·0735	·1024	·1181	·1672	·2128	·2408	·2461
	6	·0000	·0000	·0001	·0006	·0028	·0087	·0210	·0341	·0424	·0743	·1160	·1542	·1641
	7	·0000	·0000	·0000	·0000	·0003	·0012	·0039	·0073	·0098	·0212	·0407	·0635	·0703
	8	·0000	·0000	·0000	·0000	·0000	·0001	·0004	·0009	·0013	·0035	·0083	·0153	·0176
	9	·0000	·0000	·0000	·0000	·0000	·0000	·0000	·0001	·0001	·0003	·0008	·0016	·0020
10	0	·9044	·5987	·3487	·1969	·1074	·0563	·0282	·0173	·0135	·0060	·0025	·0012	·0010
	1	·0914	·3151	·3874	·3474	·2684	·1877	·1211	·0867	·0725	·0403	·0207	·0114	·0098
	2	·0042	·0746	·1937	·2759	·3020	·2816	·2335	·1951	·1757	·1209	·0763	·0495	·0439
	3	·0001	·0105	·0574	·1298	·2013	·2503	·2668	·2601	·2522	·2150	·1665	·1267	·1172
	4	·0000	·0010	·0112	·0401	·0881	·1460	·2001	·2276	·2377	·2508	·2384	·2130	·2051
	5	·0000	·0001	·0015	·0085	·0264	·0584	·1029	·1366	·1536	·2007	·2340	·2456	·2461
	6	·0000	·0000	·0001	·0012	·0055	·0162	·0368	·0569	·0689	·1115	·1596	·1966	·2051
	7	·0000	·0000	·0000	·0001	·0008	·0031	·0090	·0163	·0212	·0423	·0746	·1080	·1172
	8	·0000	·0000	·0000	·0000	·0001	·0004	·0014	·0030	·0043	·0106	·0229	·0389	·0439
	9	·0000	·0000	·0000	·0000	·0000	·0000	·0001	·0003	·0005	·0016	·0042	·0083	·0098
	10	·0000	·0000	·0000	·0000	·0000	·0000	·0000	·0000	·0000	·0001	·0003	·0008	·0010

APPENDIX II

Bionomial Coefficients

An entry in the table gives the value of $\underline{|n}/\underline{|x}\ \underline{|n-x}$ for $n=2, 3, \ldots\ldots, 20$ and $x=2, 3, \ldots\ldots, 10$.

$n \backslash x$	2	3	4	5	6	7	8	9	10
2	1								
3	3	1							
4	6	4	1						
5	10	10	5	1					
6	15	20	15	6	1				
7	21	35	35	21	7	1			
8	28	56	70	56	28	8	1		
9	36	84	126	126	84	36	9	1	
10	45	120	210	252	210	120	45	10	1
11	55	165	330	462	462	330	165	55	11
12	6	220	495	792	924	792	495	220	66
13	78	286	715	1,287	1,716	1,716	1,287	715	286
14	91	364	1,001	2,002	3,003	3,432	3,003	2,002	1,001
15	105	455	1,365	3,003	5,005	6,435	6,435	5,005	3,003
16	120	560	1,820	4,368	8,008	11,440	12,870	11,440	8,008
17	136	680	2,380	6,188	12,376	19,448	24,310	24,310	19,448
18	153	816	3,060	8,568	18,564	31,824	43,758	48,620	43,758
19	171	969	3,876	11,628	27,132	50,388	75,582	92,378	92,378
20	190	1,140	4,845	15,504	38.760	77.520	125,970	167,960	184,756

ABOUT THE BOOK

1. Strictly according to the new C.B.S.E. syllabus.
2. Simple yet rigorous exposition.
3. (i) Contains short biographical notes and portraits of famous mathematicians.

   (ii) Graded exercises, summary, review exercise and historical notes in each chapter.

   (iii) Selected questions set at CBSE, IITJEE, AIIMS, Roorkee Ent. Tests, etc.

\# \# \# \# \# \# \# \# \#

ABOUT THE AUTHORS

Dr. M. K. SINGAL is Professor and Head of the Department of Mathematics at Meerut University. A Fellow of the National Academy of Sciences (India) and the Institute of Mathematics and its Applications (U.K.), he has been president of the Indian Mathematical Society. He has been associated with NCERT and UGC for improvement of mathematical education in the country. He is the author of over 50 books at school and college level.

Dr. (Mrs.) ASHA RANI SINGAL is Professor of Mathematics at Meerut University. She has been associated with preparation of instructional materials for NCERT, Indira Gandhi National Open University, and UGC. She is a Fellow of the National Academy of Sciences (India), and Institute of Mathematics and its Applications (U.K.), etc. She is the author of several books.